

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



**The association between perfectionism and Body Dysmorphic Disorder (BDD)  
a systematic review**

Quinn, Rachel Elizabeth

*Awarding institution:*  
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

**END USER LICENCE AGREEMENT**



**Unless another licence is stated on the immediately following page** this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

**Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# VOLUME I

## *Systematic Review & Main Empirical Project*

**Rachel Quinn**

May 2018

**Quinn, Rachel**

Thesis submitted in partial fulfilment of the degree of Doctorate in Clinical Psychology

## ACKNOWLEDGEMENTS

---

This thesis would not have been possible without the continuous support from my research supervisors, Dr Amita Jassi and Dr Georgina Krebs. Thank you for always going beyond what I ever expected of you. Your support, wisdom, and expertise were invaluable, and I have gained so much from you. My clinical tutor, Dr Kate Johnson, also deserves special thanks for helping me to see the light at the end of the tunnel and for always knowing the right thing to say.

Thank you to the school for allowing me to carry out my research there. You were so generous with your time in assisting me to organise the study, and this project would not have worked without you. A special thanks also goes to the pupils who gave their time in completing the questionnaire.

I would not have been able to get through the last three years without my extremely supportive cohort. It has been a privilege to not only work alongside you, but to share this experience and develop some friendships that I will value for the rest of my life. Despite the stress and pressure of writing up this work, your understanding and humour have made even the ASB an enjoyable place to be.

And finally, I would like to thank my partner, family, and friends for knowing exactly how to make me smile through the tough bits, and for continuously supporting (and tolerating!) me on this challenging but worthwhile journey.

# TABLE OF CONTENTS

---

## SYSTEMATIC REVIEW

### **The association between perfectionism and Body Dysmorphic Disorder (BDD):**

A systematic review

*Page 4*

## MAIN EMPIRICAL PROJECT

### **Understanding “normal” appearance concerns in young people and how they are related to perfectionism:** A cross-sectional study in a British school sample

*Page 88*

# SYSTEMATIC REVIEW

---

## **The association between perfectionism and Body Dysmorphic Disorder (BDD):** *A systematic review*

---

**First Supervisor:** Dr Amita Jassi

**Second Supervisor:** Dr Georgina Krebs

## ABSTRACT

---

Perfectionism is a transdiagnostic process associated with a variety of clinical disorders. Although perfectionism has been postulated as being involved in Body Dysmorphic Disorder (BDD), it is unclear whether this association is supported by empirical data and no previous systematic review has attempted to synthesis findings in this field. The aim of the current review was therefore to consolidate published literature that has explored the relationship between perfectionism and BDD. Given the limited literature on BDD patients, this review included studies of the association between perfectionism and appearance-related concerns in analogue samples, unselected samples, and other clinical groups. However, studies that exclusively focussed on body weight/shape concerns pertaining to eating disorders were excluded. Three electronic databases were searched: PsychINFO, PubMed, and Web of Science, returning a total of 1,047 articles, where 41 published articles met the inclusion criteria, utilising 44 independent samples. Only two studies included clinical samples of patients with BDD. The outcome measures used to assess perfectionism and appearance-related concerns were diverse. Nevertheless, 90% of cross-sectional studies reported significant associations between perfectionism and appearance-related concerns. Therefore, in the context of largely cross-sectional data, perfectionism does appear to be reliably associated with appearance-related concerns. The review also demonstrated that Concern over Mistakes (CoM), Doubts about Actions (DaA), Socially-Prescribed Perfectionism (SPP), and maladaptive dimensions of perfectionism were the most consistently related to appearance-related outcomes. These findings need to be considered alongside their limitations. In particular, the vast majority of studies were conducted in unselected student samples, and so it remains unclear whether these findings generalise to patients diagnosed with BDD. Notwithstanding the limitations, the current review highlights the potential role of perfectionism in BDD and the need for further research in this field.

(PROSPERO ID = CRD42017055874)

# CONTENTS

---

<b>1. INTRODUCTION .....</b>	<b>8</b>
1.1. BDD and Appearance-related Concern .....	8
1.2. Perfectionism.....	9
1.3. Perfectionism and BDD.....	11
1.4. The Current Review .....	12
<b>2. METHOD.....</b>	<b>13</b>
2.1. Procedure .....	13
2.1.1. Search Strategy .....	13
2.1.2. Eligibility Criteria .....	13
2.2. Data Extraction .....	14
2.3. Quality Assessment.....	14
<b>3. RESULTS .....</b>	<b>13</b>
3.1. Overview of Studies .....	16
3.2. Study Characteristics .....	16
3.3. Measurement .....	17
3.3.1. Perfectionism measures .....	17
3.3.2. Appearance-Related Concern Measures .....	21
3.4. Quality Assessment of Included Studies.....	21
3.4.1. Longitudinal/Cross-sectional studies .....	22
3.4.2. Case-Control studies .....	22
3.5. Main Findings.....	26
3.5.1. Is there an association between perfectionism & appearance-related concerns? .....	26
3.5.2. Are some dimensions of perfectionism more strongly or consistently associated with appearance-related concerns compared to others? .....	26
3.5.3. Does the association still exist when controlling for other mental health diagnoses? .....	45
3.5.4. Can any causal conclusions be drawn? .....	45
3.5.5. Do the associations vary depending on primary diagnosis?.....	46
<b>4. DISCUSSION.....</b>	<b>47</b>
4.1. Summary.....	47
4.1.1. Is there an association between perfectionism and BDD? .....	48
4.1.2. Are some dimensions of perfectionism more strongly or consistently associated with appearance-related concerns? .....	48
4.1.3. Does the association still exist when controlling for other mental health diagnoses? .....	50

4.1.4. Can any causal conclusions be drawn? .....	50
4.1.5. Do the associations vary depending on primary diagnosis?.....	50
4.2. General Limitations.....	51
4.3. Implications & Future Directions .....	52
4.4. Conclusions .....	53
<b>5. REFERENCES .....</b>	<b>55</b>
<b>6. APPENDIX .....</b>	<b>66</b>
6.1. Table summarising the perfectionism assessment measures used by the studies. ..	66
6.2. Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies .....	69
6.3. Quality Assessment Tool for Case-Control Studies.....	77
6.4. Table summarising appearance-related outcome measures used by the studies ....	84



# 1. INTRODUCTION

---

## 1.1. BDD and Appearance-related Concerns

Body Dysmorphic Disorder (BDD) is characterised by appearance-related concerns, where these clinical concerns differ both qualitatively and quantitatively from normative appearance-related concerns expressed frequently by the general population (Veale, 2004). Those with the diagnosis demonstrate a preoccupation with “defected” body feature(s) that cause significant distress and impairment in functioning, although they are not perceivable to others (American Psychiatric Association, 2013). Evidence suggests that references to BDD date back to 1891, and therefore it is not simply a construct of modern society as some might expect (Veale & Neziroglu, 2010). Despite being evident throughout psychiatric history, it has only been in the last two decades that scientific research on BDD has emerged (Jerome, 2017), where in the American literature, a 1991 review only identified five published articles (Phillips, 1991). BDD was not included in the “Diagnostic and Statistical Manual of Mental Disorders” (DSM) until its third edition in 1980, and since this inclusion it has gained more momentum and undergone meaningful changes to be considered an independent disorder (Phillips, 2017). Due to this much overdue evidence-base, a lack of awareness still exists clinically with BDD comparative to other disorders, where BDD often goes undetected or misdiagnosed in clinic (Veale, Akyüz, & Hodson, 2015). Subsequently, there is still a large amount of stigma surrounding the disorder (Veale & Neziroglu, 2010), and reasons for not disclosing symptoms include: embarrassment; fear of negative judgement; not expecting clinicians to understand; unawareness that treatment exists; not feeling that the symptoms were a big enough problem; not wanting to know they were problematic; and not knowing that others had similar issues (Conroy et al., 2008).

Several cognitive-behavioural models of the development and maintenance of BDD exist (Feusner, Neziroglu, Wilhelm, Mancusi & Bohon, 2010; Neziroglu, Khemlani-Patel & Veale, 2008; Veale, 2004; Wilhelm & Neziroglu, 2002; Wilhelm, Phillips, & Steketee, 2013). These models collectively suggest that individuals with BDD experience negative thoughts and emotions associated with their appearance, and so partake in self-defeating coping strategies aimed at reducing appearance-related distress, including repetitive rituals and avoidance (Rasmussen, Gómez & Wilhelm, 2017). Further, dissatisfaction with appearance is considered a common experience within the models, but the individuals who go on to develop BDD engage in several maladaptive cognitive and emotional processes in response to these “normative” concerns (such as: dysfunctional

appearance-related beliefs, interpretive biases; selective attention; enhanced aesthetic perceptual sensitivity; and emotion regulation) (Buhlmann & Hartmann, 2017). These in turn exaggerate their perception, resulting in the individual processing themselves as an aesthetic object (Veale, 2004). In addition, it is hypothesised that life experience (e.g. culture, bullying, societal pressures, and family/personal history), biological factors (e.g. brain chemistry, genetics), and psychological factors (e.g. personality) can increase a person's vulnerability to developing BDD, where a key personality trait implicated in this is perfectionism (Wilhelm et al., 2013).

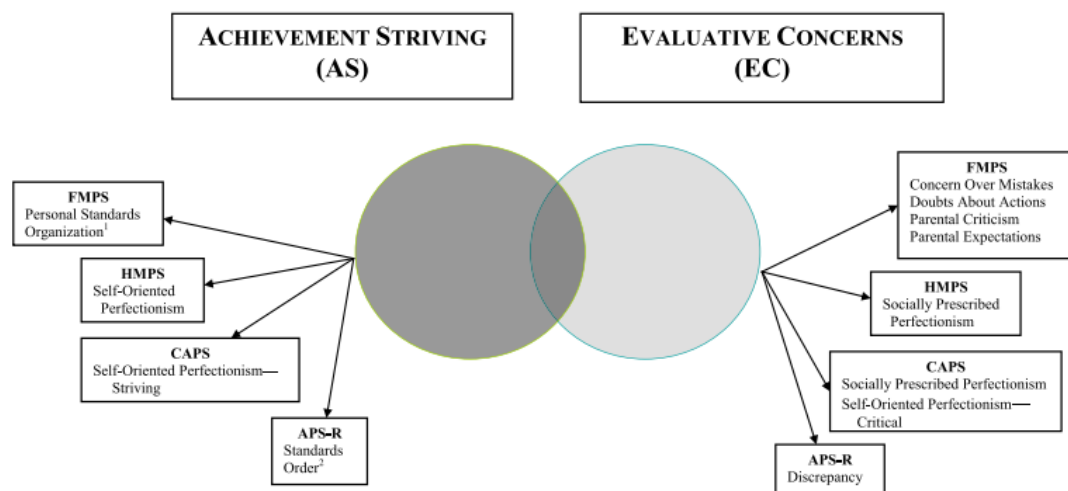
## 1.2. Perfectionism

Perfectionism has been described as the setting of excessively high standards for performance, paired with an overly critical self-evaluation and fear of failure (Frost, Marten, Lahart & Rosenblate, 1990; Hewitt & Flett, 1991). In the early 1990s, consistent clinical descriptions and observations led to the view that perfectionism had more than one dimension. For example, it was shown to include various behaviours such as concern over mistakes, doubting, standard setting, as well as interpersonal processes (e.g. pressure from the self and/or others) (Shafran, Cooper & Fairburn, 2002). This subsequently led to the development of various tools to assess these dimensions, the most distinguished being the Frost Multidimensional Perfectionism Scale (F-MPS; Frost et al., 1990) and the Hewitt and Flett Multidimensional Perfectionism Scale (HF-MPS; Hewitt & Flett, 1991).

The F-MPS is an expanded measure of perfectionism consisting of 35-items, and the six dimensions were derived from the literature and tested by the authors. The six dimensions are: Personal Standards (PS); Concerns over Mistakes (CoM); Organisation (O); Doubts about Actions (DaA); Parental Expectations (PE); and Parental Criticism (PC). The HF-MPS includes 45-items across three dimensions. That is, 1) Self-Oriented Perfectionism (SOP), 2) Socially-Prescribed Perfectionism (SPP), and 3) Other-Oriented Perfectionism (OOP). Although since its development, it is common within the literature to exclude OOP, as it has frequently been observed to lack relationships with variables and not hold any unique correlations (Frost, Heimberg, Holt, Mattia, & Neubauer, 1993; Stoeber, 2014). This tool was further adapted for children and adolescents, named the Child and Adolescent Perfectionism Scale (CAPS; Flett et al., 2016). It has 22-items and the same two SOP and SPP dimensions, although OOP was not included due to the lack of developmental evidence for this construct in young people. Because of the already well-established relationship between perfectionism and Eating Disorders (EDs) and Obsessive-Compulsive Disorder (OCD), such diagnostic assessment tools can also measure perfectionism, however these tend to be

unidimensional and embedded as a subscale within the diagnostic tool. Examples include the Eating Disorders Inventory (EDI; Garner, 1991) and the Obsessive Beliefs Questionnaire (OBQ; Obsessive Compulsive Cognitions Working Group, 2005).

The development of such tools allowed for the accumulation of evidence illustrating the association between perfectionism and various mental health diagnoses, although more recently, a growing consensus has emerged favouring a bi-dimensional model of perfectionism (DiBartolo and Rendón, 2012). That is, 1) “achievement striving” (also termed “adaptive”, “positive”, or “high standards”) and 2) “evaluative concerns” (also termed “maladaptive”, “negative”, or “neurotic”). This is because people can hold ambivalent attitudes towards perfectionism, as it is not uniquely dysfunctional but can be relatively “healthy” by encouraging positive striving (Shafran & Mansell, 2001; Slaney, Rice, Mobley, Trippi, & Ashby, 2001). In response, Slaney and colleagues (2001) developed the Almost Perfect Scale - Revised (APS-R), which measures three facets of perfectionism that discriminate between adaptive and maladaptive perfectionism, namely: 1) Standards, 2) Order, and 3) Discrepancy (i.e. maladaptive) (Slaney, 2015). Previously established tools have also undergone subsequent factor analysis to evaluate how well their dimensions load onto this bi-dimensional model, where the CoM, DaA, PC, and PE dimensions of the F-MPS and SPP of the HF-MPS are thought to represent maladaptive perfectionism, and PS and O of the F-MPS and SOP of the HF-MPS are considered adaptive (DiBartolo and Rendón, 2012). Please see Figure 1 for a visual representation of this.



**Fig. 1.** The bi-dimensional model of perfectionism<sup>1</sup> (Description of illustration: two interlocking circles each labeled with appropriate subscales). *Notes.* FMPS = Frost et al. Multidimensional Perfectionism Scale; HMPS = Hewitt & Flett Multidimensional Perfectionism Scale; CAPS = Child and Adolescent Perfectionism Scale; APS-R = Almost Perfect Scale-Revised. Based on factor analyses conducted by Frost et al. (1993); Bieling et al. (2004); Cox et al. (2002); Dunkley et al. (2003); and Pearson and Gleaves (2006). <sup>1</sup>For specific definitions of each subscale, see Flett and Hewitt (2002a). <sup>2</sup>Pearson and Gleaves (2006) found a third perfectionism factor that consisted of the FMPS Organization and APS-R Order subscales.

**Figure 1: Visual representation of how various measures of perfectionism load on to the bi-dimensional model, cited from DiBartolo and Rendón (2012, p. 141).**

Similarly, many other perfectionism tools exist which reflect various structures and dimensions. Those that will be observed later include the Perfectionistic Self-Presentation Scale (PSPS; Hewitt et al., 2003), the Positive and Negative Perfectionism Scale (PANPS; Terry-Short, Owens, Slade, & Dewey, 1995), the Neurotic Perfectionism Questionnaire (NPQ; Mitzman, Slade, & Dewey, 1994) and the Perfectionism Inventory – Need for Approval subscale (Hill et al., 2004). Appendix 6.1. provides a comprehensive table of each measure of perfectionism used by the studies in this review, accompanied by definitions of each dimension.

### 1.3. Perfectionism and BDD

Perfectionism has been found to be elevated across numerous anxiety disorders, depression, and EDs compared to healthy controls, and has therefore been conceptualised as a “transdiagnostic process” (i.e. contributing to the aetiology and maintenance of multiple disorders; for review, see Egan, Wade, & Shafran, 2011). Extensive scientific reviews have also previously demonstrated the robust relationship specifically between perfectionism and EDs, as well as OCD (Bardone-Cone et al., 2007; Egan et al., 2011; Pinto et al., 2017; Sassaroli et al., 2008). BDD shares properties with both these disorders, whereby EDs are similarly characterised by appearance-related concerns (Veale & Neziroglu, 2010), and obsessions and ritualistic behaviour are imperative features of both BDD and OCD - hence it’s reclassification in the latest edition of the DSM as a “related disorder” to OCD (Jerome, 2017). However, due to the issues discussed earlier, BDD has received significantly less attention comparative to EDs and OCD, where even reviews of general psychopathology and perfectionism have not been inclusive of BDD. It is not surprising then, that there is yet to be a systematic review consolidating the relationship between BDD and perfectionism.

Given the limited literature on BDD more generally, it was anticipated that there would be very few studies conducted examining the association between perfectionism and BDD in clinical samples. Therefore, it was deemed appropriate to widen our search and include studies investigating healthy populations and/or which used analogue samples. It was further deemed appropriate to accept studies that measured BDD-associated constructs, such as general appearance-related concerns. However, to preserve the review’s focus on BDD, it was imperative that any appearance-related measure focusing on weight and/or shape concerns was excluded as these constructs would be more representative of EDs.

## 1.4. The Current Review

Independent research has begun to emerge indicating an association between BDD and perfectionism, although a summary of the findings is warranted. By collating and assessing the quality of the evidence, this review would be the first to contribute a collective understanding of the role of perfectionism in BDD, allowing for specific recommendations to be made about how this can be considered clinically, as well as where research in this field should be directed. Given that research on BDD is in its infancy, it was anticipated that there would be relatively few studies of perfectionism in BDD patients. We therefore included studies of the association between perfectionism and general appearance-related concerns (e.g. body dissatisfaction, appearance anxiety) measured in non-BDD populations, such as analogue or unselected samples. However, studies that exclusively focussed on body weight/shape concerns pertaining to eating disorders were excluded.

### Primary research question:

- Is there an association between perfectionism and BDD?

### Secondary research questions:

- Are some dimensions of perfectionism more strongly or consistently associated with BDD compared to others?
- Does the association between perfectionism and BDD still exist when controlling for other mental health diagnoses, including anxiety and depression?
- Can any causal conclusions be drawn from the literature?
- Do the associations between perfectionism and BDD vary depending on primary diagnosis (ED versus OCD versus BDD)?

## 2. METHOD

---

### 2.1. Procedure

The PRISMA guidelines (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009) were used to inform the methods of this systematic review. It was also registered on PROSPERO, the international prospective register of systematic reviews (ID = CRD42017055874).

#### 2.1.1. Search Strategy

The search was conducted on 9<sup>th</sup> June 2017. Articles were identified by searching three electronic databases: PsychINFO, PubMed, and Web of Science. To identify the studies that were only relevant to the current research question, specific search terms were identified and adapted according to each database's search requirements. The search criteria were: "body image" OR "body concern\*" OR "appearance concern\*" OR "body dissatisfaction" OR "appearance dissatisfaction" OR "body satisfaction" OR "appearance satisfaction" OR "body dysmor\*" OR "dysmorphophobia" OR "appearance anxiety" OR "appearance worr\*" OR "body dysmorphic disorder" AND "perfection\*". Asterisks represent a truncation instruction. These search terms were also reviewed by two other researchers to ensure that the terms were comprehensive enough to yield the maximum number of articles. The titles and abstracts of articles were searched without any restrictions or filters to determine if these words were present.

Figure 2 in the 'Results' section demonstrates the process used to select the final studies. That is, after duplicated articles were removed, all titles and abstracts were screened. Abstracts referring to the variables of interest were included, and the full-texts were acquired and assessed for eligibility. If it was not clear from the abstract whether the study met the eligibility criteria, the full-text was reviewed. If after reviewing the full-text and the eligibility was still uncertain, other researchers were consulted. Where full-text access was not available, the first authors were contacted to request these.

#### 2.1.2. Eligibility Criteria

For an article to be included in this review, it had to examine the association between perfectionism and BDD or associated appearance-related concerns, and be a peer-reviewed, English article with quantitative data. All study designs, samples, and publication years were accepted. Studies were

therefore excluded if they: were not published in English; had solely qualitative data; were not peer-reviewed (i.e. book chapter, dissertation); or were a case study or review article. In addition, studies were required to have measured perfectionism and BDD/appearance-related concerns using independent assessment tools (i.e. they were not sub-scales of the same measure, such as the “Body Dissatisfaction” and “Perfectionism” subscales from the EDI). The measure of appearance-related concern had to capture general body dissatisfaction or appearance anxiety, as this was considered to be closely related to BDD symptoms. Studies that exclusively focussed on body weight/shape concerns pertaining to eating disorders were excluded.

To evaluate the association between perfectionism and BDD, it was essential that the studies used appropriate statistical analysis and reported on the direct association between these two variables (e.g. correlation or regression).

## 2.2. Data Extraction

To appropriately review the eligible studies, several pieces of data/information were extracted from each article and recorded in an Excel spreadsheet. This included: the first author; date of publication; study design; sample size, age, gender, and setting; diagnostic status; measure of BDD/appearance-related concern; measure of perfectionism; dimensions of perfectionism measured; statistical test used and statistical co-efficient reported; any covariate variables; and the results of the analysis. The Excel data extraction tool was developed independently, and if articles did not report any of the required information then this was coded as NR (“not reported”).

## 2.3. Quality Assessment

To assess the internal validity and individual bias present within each of the eligible studies, a quality assessment was undertaken. The National Institute of Health (NIH) have developed a range of quality assessment tools for studies of various designs. An initial literature review indicated that most of the studies in this area were cross-sectional, with some also being longitudinal in design, and even fewer comparing cases and controls. The current review therefore utilised the NIH’s tools for each of these study designs. The tool and accompanying guidance can be found in Appendix 6.2. and 6.3. These tools were not designed to provide a list of factors and a numeric quality score, but rather assist the researcher to critically appraise the study by drawing their attention to various key concepts (NIH, 2014). Twenty per cent of the eligible studies were reviewed by a second quality

assessor to enhance reliability. Those studies where the two raters were not in agreement sought the consultation of a third researcher.



### 3. RESULTS

#### 3.1. Overview of Studies

As demonstrated in Figure 2, searches on the three databases yielded a total of 1,047 published research papers. After removing duplicates, the titles and abstracts of 709 articles were screened and the full-texts were accessed if they referred to the variables of interest ( $N = 151$ ). Full-texts were assessed against inclusion/exclusion criteria to determine their eligibility for the review, where 41 studies were accepted in the final sample. Moreover, three of these studies (Barnett & Sharp, 2016; Levinson et al., 2013; Sherry et al., 2009) reported on data from two separate samples. For the purposes of this review, these were deemed to be independent studies, and therefore a total of 44 studies were reported on.

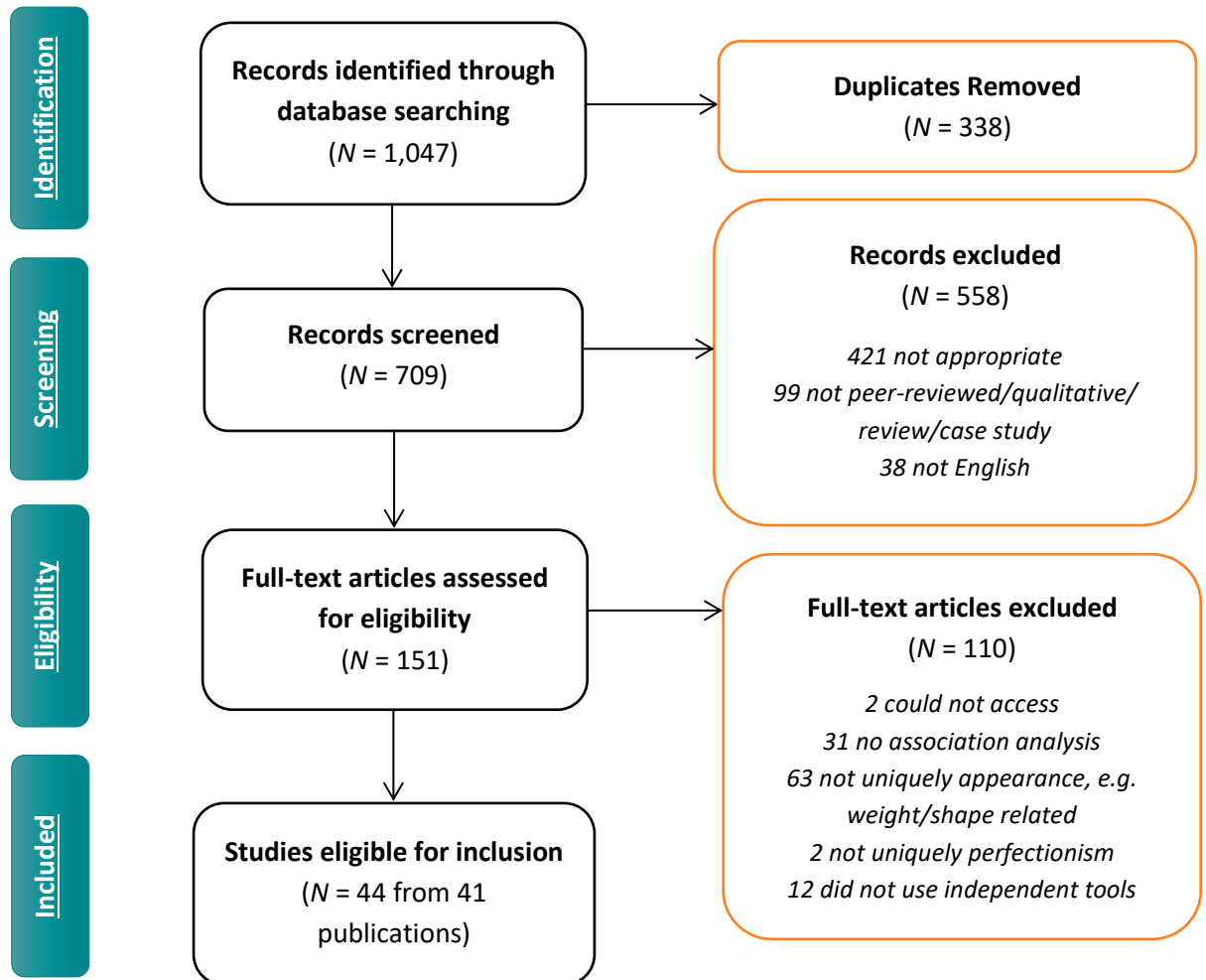


Figure 2: PRISMA Diagram of the search procedure

## 3.2. Study Characteristics

Table 1 presents the characteristics of the 44 included studies. Studies dated from 1995 up until 2017, and were predominantly cross-sectional in design (89%,  $N = 39$ ). Only two studies (4.5%) included patients with clinical diagnoses of BDD (Buhlmann, Etcoff, & Wilhelm, 2008; Hartmann, Thomas, Greenberg, Matheny & Wilhelm, 2014). Two further studies (4.5%) included patient samples but were patients with eating disorders diagnoses (Davis, 1997) or overweight/obese participants (Malkina-Pykh, 2012). The remaining 40 studies (91%) used unselected, healthy participants recruited from largely student settings. Of note, two of these studies included additional analysis using an analogue sub-sample of participants scoring highly on a self-report measure for BDD (Bartsch, 2007; Schieber, Kollei, de Zwaan, Müller & Martin, 2013). The studies were conducted across 10 different countries – although these were mostly from the USA (55%,  $N = 24$ ), Australia (16%,  $N = 7$ ), and Canada (11%,  $N = 5$ ). There was large variation in total sample sizes, which ranged from 33 to 2,071. However, a significant proportion of studies recruited uniquely from student populations (64%,  $N = 28$ ), with only five studies recruiting from schools with samples below the age of 18 (11%) – although an additional two studies had samples with a mean age below 18 years but were also inclusive of adults. Regarding the gender of the samples, most studies recruited exclusively from female populations (45%,  $N = 20$ ), or had a predominantly female sample (20%,  $N = 9$ ). Seven studies (16%) included only male participants, and approximately 18% ( $N = 8$ ) had a relatively balanced sample (i.e. between 45-55% of male participants). Interestingly, a third of the studies did not report on their sample's ethnic composition (34%,  $N = 15$ ). Of the information that was available, the ethnic diversity of the samples was slight, with only three studies (7%) having samples where less than half identified as being “white” or similar (“White/European”, “Caucasian”, or “Australian”). Just under a third (32%,  $N = 14$ ) reported a relatively homogeneous sample (with over 75% identifying as “white”).

## 3.3. Measurement

### 3.3.1. Perfectionism measures

In total, perfectionism was assessed using 10 different measures (see Appendix 6.1. for details). Some included different dimensions of perfectionism and some were unidimensional. Tables 2 – 6 summarise the findings from the 44 studies and are grouped according to the measure of perfectionism used. Four papers included two independent measures of perfectionism, and therefore appear twice within the tables.

**Table 1: Summary of the study characteristics (N = 44)**

1 <sup>st</sup> Author and Year						Design	N	Diagnostic Status	Country	Setting	M age (Years)	% Male	% White*
Bardone-Cone (2008)		Cross-sectional	347	Healthy	USA	University Students	19	40	87				
Bardone-Cone (2017)		Cross-sectional	441	Healthy	USA	University Students	19	0	69				
Barnes (2017)		Cross-sectional	220	Healthy	Australia	University Students and Community	24	21	NR				
Barnett (2016)	STUDY 1	Cross-sectional	580	Healthy	USA	University Students	NR	0	53				
	STUDY 2	Cross-sectional	398	Healthy	USA	University Students	NR	0	44				
Bartsch (2007)		Cross-sectional	619 14	Healthy group Probable BDD sub-group	Australia	University Students	26 25	27 14	76 86				
Blakey (2016)		Cross-sectional	601	Healthy	USA	University Students	20	45	72				
Brannan (2008)		Cross-sectional	398	Healthy	USA	University Students	20	0	65				
Brannan (2009)		Cross-sectional	204	Healthy	USA	Collegiate Athletes	20	0	76				
Brosof (2017)		Longitudinal	300	Healthy	USA	University Students	18 (Median)	0	61				
Buhlmann (2008)		Case-Control	19 21 21	BDD group OCD group Healthy Control	USA	OCD Clinic Patients OCD Clinic Patients Community	33 32 34	32 48 43	100 100 100				
Cash (1995)		Cross-sectional	284	Healthy	USA	University Students	22	0	74				
Chen (2010)		Cross-sectional	883	Healthy	Taiwan	Junior High School Students (12-16 years)	NR	51	NR				
Cooley (2007)		Cross-sectional	339	Healthy	USA	University Students	18	0	90				
Custers (2009)		Cross-sectional	711	Healthy	Belgium	Secondary School Students	14	45	NR				
Dakanalis (2015 <sub>a</sub> )		Cross-sectional	551	Healthy	Italy	University Students	21	100	95				

1 <sup>st</sup> Author and Year	Design	N	Diagnostic Status	Country	Setting	M age (Years)	% Male	% White*
Dakanalis (2014)	Cross-sectional	605	Healthy	Italy	University Students	21	100	95
Davis (1997)	Cross-sectional	123	AN, BN, EDNOS group	Canada	Hospital Patients	28	0	NR
Davis (2005)	Cross-sectional	100	Healthy	Canada	University Students	23	100	NR
Dour (2011)	Cross-sectional	161	Healthy	USA	Middle School Students	13	46	84
Dryer (2016)	Cross-sectional	158	Healthy	Australia	Community	27	100	NR
Dunn (2011)	Cross-sectional	119	Healthy	Canada	Competitive Figure Skaters	15	0	NR
Ferrand (2007)	Cross-sectional	33	Healthy	France	Synchronized Swimmers (Division 1)	17	0	NR
Frederick (2016)	Cross-sectional	488	Healthy	USA	University Students	20	0	57
Grammas (2009)	Cross-sectional	202	Healthy	USA	University Students	22	100	28
Hanstock (2002)	Cross-sectional	165	Healthy	Australia	University Students	20	0	NR
Hartmann (2014)	Case-Control	23	BDD group	USA	All groups: Outpatients clinics/residential treatment settings/community.	30	26	78
		24	AN group			26	8	88
		22	Healthy Control			29	32	73
Lamanna (2010)	Cross-sectional	348	Healthy	USA	University Students	19	29	87
Lavell (2014)	Cross-sectional	246	Healthy	Australia	University Students	21	26	NR
Levinson (2016)	Longitudinal	300	Healthy	USA	University Students	18 (Median)	0	61
Levinson (2013) STUDY 1	Cross-sectional	236	Healthy	USA	University Students	19 (Median)	26	90
	STUDY 2	156	Healthy	USA	University Students	19 (Median)	0	46
Malkina-Pykh (2012)	Cross-sectional	104	Overweight/obese group	Russia	Attendees of a weight management programme	38	31	NR

1 <sup>st</sup> Author and Year	Design	N	Diagnostic Status	Country	Setting	M age (Years)	% Male	% White*
Menatti (2013)	Cross-sectional	167	Healthy	USA	University Students	19	0	90
Midlarsky (2008)	Cross-sectional	290	Healthy	USA	Community	52	0	92
Minnich (2014)	Cross-sectional	302	Healthy	USA	University Students	19	100	89
Murray (2013)	Cross-sectional	119	Healthy	Australia	University Students	22	100	NR
Rasooli (2011)	Cross-sectional	373	Healthy	Iran	University Students	NR	0	NR
Rudiger (2007)	Cross-sectional	121	Healthy	USA	University Students	21	0	68
Saling (2005)	Longitudinal	326	Healthy	Australia	Primary School Students	9	54	NR
Schieber (2013)	Cross-sectional	2,071 58	Healthy group Probable BDD sub-group	Germany	Community	45	47	NR
						43	19	NR
Shaw (2004)	Cross-sectional	496	Healthy	USA	Middle School Students	13	0	68
Sherry (2009)	STUDY 1	96	Healthy	Canada	Community members to a fitness facility	28	54	60
	STUDY 2	118	Healthy	Canada	University Student members to a fitness facility	23	53	58

*Note:* NR = Not Reported; BDD = Body Dysmorphic Disorder; OCD = Obsessive Compulsive Disorder; AN = Anorexia Nervosa; BN = Bulimia Nervosa; EDNOS = Eating Disorders Not Otherwise Specified. White\* = participants described as “White”, “Caucasian”, “White/European”, “Australian”.

Two measures of perfectionism were utilised particularly frequently. The first being the F-MPS, used by 14 of the studies (29%), where each study included a different combination of dimensions, although the maladaptive/high standards classification was used by four studies (both samples in Levinson et al., 2013; Levinson & Rodebaugh, 2016; Menatti, Weeks, Levinson, & McGowan, 2013). The other most prevalent tool was the HF-MPS, which was equally utilised by 29% ( $N = 14$ ) of studies. Almost half of the 14 papers assess all three dimensions of perfectionism; however, a larger proportion did not consider Other-Oriented Perfectionism (OOP).

The EDI was used by seven studies (15%), 8% ( $N = 4$ ) used the APS-R, and the remaining nine studies utilised a variety of perfectionism tools. This included the PSPS (6%,  $N = 3$ ), the OBQ-PC (4%,  $N = 2$ ), the PANPS (2%,  $N = 1$ ), the CAPS (2%,  $N = 1$ ), the NPQ (2%,  $N = 1$ ), and the Perfectionism Inventory - Need for Approval subscale (2%,  $N = 1$ ).

### 3.3.2. Appearance-Related Concern Measures

Thirty-two different measures of appearance-related concern were used across the 44 studies. These are listed in Appendix 6.4. with a description and reference. Thirteen measures were used in more than one study, with the maximum occurrence for one specific measure being four. Both the Social Appearance Anxiety Scale (Hart et al., 2008) and the Appearance subscales of the Multidimensional Body-Self Relations Questionnaire (MBSRQ; Brown, Cash, Mikulka, 1990) were used by four studies each. Generally, these measures assessed appearance-related: esteem; symptoms (of body/muscle dysmorphia); anxiety/distress; behaviours (e.g. body checking, motivation to increase muscularity); and thought processes (e.g. dysmorphic concerns, cognitive distortions, schemas, beliefs, attitudes, satisfaction).

## 3.4. Quality Assessment of Included Studies

The quality of each study was critically reviewed and overall was judged as having “good”, “fair”, or “poor” quality. The NIH (2014) advise that a “good” study has the least risk of bias and results are considered valid; a “fair” study is susceptible to some bias, but this is not enough to invalidate the results; and a “poor” study is one with significant risk of bias and the results are questionable. There was 100% agreement between the two raters on the presence/absence of quality criteria.

#### 3.4.1. Longitudinal/Cross-sectional studies

Forty-two of the 44 studies had either a cross-sectional or longitudinal design. These were subsequently reviewed using the tool and guidance presented in Appendix 6.2. Table 2 presents the overall quality ratings given to the 42 studies, as well as their rating for each quality criterion. Studies were either rated as meeting this criteria (Y) or not (N), or whether this item was not applicable (NA), could not be determined (CD), or not reported (NR). Table 3 further provides a description of each of the individual quality criteria.

It was rare for studies to not have clearly defined the research question, study population, exposure and/or outcome measures. It was also highly likely that subjects were recruited from the same/similar populations, although it was noted that a large proportion of the studies failed to report the time of when subjects participated in the research, as well as the inclusion/exclusion criteria (or whether these were not applicable). Due to only three studies having a longitudinal design, the criteria regarding the loss to follow-up statistics were predominantly not relevant (although only one of the three studies achieved this), and the blinding of the researchers was not applicable for any study. Almost three quarters of the studies (74%,  $N = 31$ ) did not report the participation rate of eligible participants. For instance, they did not report how many university students had access to the self-report surveys, or how many were enrolled the cohort. Only 7% of studies ( $N = 3$ ) measured the exposure prior to the outcome and justified the power of the findings. Two of the three longitudinal studies had a sufficient time-period to reasonably expect an association between the exposure and outcome. Over half of the studies controlled for at least one “key” confounding variable (57%,  $N = 24$ ), and two (5%) examined different levels of the exposure. Overall, three studies were classified as “good” (7%), 36 as “fair” (86%), and three as “poor” (7%).

#### 3.4.2. Case-Control studies

Two of the 44 studies had a case-control design, which were both considered to have a “fair” level of quality. Both studies also clearly defined the research question and study population (cases and controls), although neither study provided a sample size justification or confirmed that the exposure occurred prior to the condition/event. It could not be determined whether either study used concurrent controls, and one study also failed to clearly define the measures used and inclusion/exclusion criteria.

**Table 2: The quality ratings given to each cross-sectional/longitudinal study (N = 42).**

Study	Quality Criteria Number (see Table 3 for details)														Overall Rating
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Bardone-Cone (2008)	Y	Y	NR	Y	Y	N	N	N	Y	N	Y	NA	NA	Y	Fair
Bardone-Cone (2017)	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	N	Fair
Barnes (2017)	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	N	Fair
Barnett (2016) STUDY 1	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	Y	Fair
STUDY 2	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	Y	Fair
Bartsch (2007)	Y	Y	Y	Y	N	N	N	Y	Y	N	Y	NA	NA	Y	Good
Blakey (2016)	Y	Y	N	Y	N	N	N	N	Y	N	Y	NA	NA	Y	Fair
Brannan (2008)	Y	Y	NR	Y	Y	N	N	N	Y	N	Y	NA	NA	N	Fair
Brannan (2009)	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	N	Fair
Brosof (2017)	Y	Y	NR	Y	N	N	N	N	Y	Y	Y	NA	N	Y	Fair
Cash (1995)	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	N	Fair
Chen (2010)	Y	Y	Y	Y	N	N	N	N	Y	N	Y	NA	NA	N	Fair
Cooley (2007)	Y	Y	NR	Y	N	CD	CD	N	Y	N	Y	NA	NA	N	Fair
Custers (2009)	Y	Y	Y	Y	N	N	N	N	Y	N	Y	NA	NA	Y	Fair
Dakanalis (2015 <sub>a</sub> )	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	Y	Fair
Dakanalis (2014)	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	Y	Fair
Davis (1997)	N	Y	NR	Y	N	N	N	N	CD	N	CD	NA	NA	Y	Poor
Davis (2005)	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	Y	Fair
Dour (2011)	Y	Y	Y	Y	N	N	N	N	Y	N	Y	NA	NA	N	Fair
Dryer (2016)	Y	Y	NR	N	N	N	N	N	Y	N	Y	NA	NA	Y	Fair



Study	Quality Criteria Number (see Table 3 for details)														Overall Rating
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Dunn (2011)	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	N	Fair
Ferrand (2007)	Y	Y	Y	Y	N	N	N	N	Y	N	Y	NA	NA	N	Fair
Frederick (2016)	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	Y	Fair
Grammas (2009)	Y	Y	NR	Y	Y	N	N	N	Y	N	Y	NA	NA	Y	Fair
Hanstock (2002)	Y	N	Y	NR	N	N	N	N	Y	N	Y	NA	NA	Y	Poor
Lamanna (2010)	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	N	Fair
Lavell (2014)	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	Y	Fair
Levinson (2016)	Y	Y	NR	Y	N	Y	Y	N	Y	Y	Y	NA	N	Y	Fair
Levinson (2013) STUDY 1	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	N	Fair
STUDY 2	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	N	Fair
Malkina-Pykh (2012)	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	N	Fair
Menatti (2013)	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	N	Fair
Midlarsky (2008)	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	N	Fair
Minnich (2014)	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	N	Fair
Murray (2013)	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	Y	Fair
Rasooli (2011)	Y	N	NR	Y	N	N	N	N	Y	N	Y	NA	NA	Y	Poor
Rudiger (2007)	Y	Y	NR	Y	N	Y	N	N	N	Y	N	NA	NA	Y	Fair
Saling (2005)	Y	Y	NR	Y	N	Y	Y	N	Y	Y	Y	NA	Y	Y	Good
Schieber (2013)	Y	Y	Y	Y	N	N	N	Y	Y	N	Y	NA	NA	Y	Good
Shaw (2004)	Y	Y	Y	Y	N	N	N	N	Y	N	Y	NA	NA	N	Fair
Sherry (2009) STUDY 1	Y	Y	N	Y	N	N	N	N	Y	N	Y	NA	NA	Y	Fair

Study	Quality Criteria Number (see Table 3 for details)														Overall Rating
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
STUDY 2	Y	Y	NR	Y	N	N	N	N	Y	N	Y	NA	NA	Y	Fair

Note: Y = Yes; N = No; NR = Not Reported; CD = Cannot Determine; NA = Not Applicable.

**Table 3: Description of each of the quality criteria with its corresponding number.**

Quality Criteria Number	Quality Criteria Description	Quality Criteria Number	Quality Criteria Description
1	Was the research question or objective in this paper clearly stated?	8	For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as continuous variable)?
2	Was the study population clearly specified and defined?	9	Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?
3	Was the participation rate of eligible persons at least 50%?	10	Was the exposure(s) assessed more than once over time?
4	Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study pre-specified and applied uniformly to all participants?	11	Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?
5	Was a sample size justification, power description, or variance and effect estimates provided?	12	Were the outcome assessors blinded to the exposure status of participants?
6	For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?	13	Was loss to follow-up after baseline 20% or less?
7	Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?	14	Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)?

## 3.5. Main Findings

### 3.5.1. Is there an association between perfectionism and appearance-related concerns?

The most commonly used statistical test to measure the association between perfectionism and appearance-related concerns was correlation analyses (see Tables 4-8). Two of these studies (5%) failed to adequately report the correlation statistic in the paper (Dour & Theran, 2011; Rasooli & Lavasani, 2011). Three studies did not use correlation (7%), where one utilised regression only, and for the remaining two this statistical approach was not applicable to the study design (case-control).

Thirty-eight out of 42 studies (90%) reported a significant association between perfectionism and appearance-related concerns, with correlation values ( $r$ ) ranging from .13 to .75. Four studies (10%) rated as “fair” quality reported no significant associations between perfectionism and appearance-related concerns (Chen, Fox, Haase & Ku, 2010; Cooley, Toray, Valdez & Tee, 2007; Custers & Van den Bulck, 2009; Shaw, Stice & Sringer, 2004). The two case-control studies with adult clinical samples from the USA were also consistent in their findings. Using Analysis of Variance (ANOVA), those diagnosed with BDD scored significantly higher compared to mentally-healthy controls on measures of perfectionism. Moreover, this was further demonstrated in the two analogue samples, where a “BDD” group were found to score significantly greater on measures of perfectionism compared to controls (using Mann-Whitney U-Tests and Analysis of Covariance [ANCOVA]).

### 3.5.2. Are some dimensions of perfectionism more strongly or consistently associated with appearance-related concerns compared to others?

To synthesise the findings from the studies in the most efficient way, this section was organised according to the measure of perfectionism used by the study:

#### F-MPS

Table 4 presents the findings of all the studies that utilised the F-MPS (all rated “fair” in quality). This table demonstrates that, even when the same measure of perfectionism is used, there is still variation in the dimensions selected to study. Nine of the 14 studies included Concern over Mistakes (CoM) as one of their dimensions, and all reported significant associations. Similarly, five studies assessed the association with Doubts about Actions (DaA), and all found significant effects.

**Table 4: Summary of the study findings from those using the Frost Multidimension Perfectionism Scale ( $N = 14$ ).**

1 <sup>st</sup> Author and Year	Measure of Appearance-Related Concerns	Measure of Perfectionism	Dimensions of Perfectionism	Statistical Analysis	Significant Findings ( $p < .05$ )*	Other Covariates
Bardone-Cone (2017)	Contingencies of Self-Worth Scale (Appearance Subscale)	F-MPS	- CoM	Correlation.	Dependence on appearance self-worth was associated with CoM ( $r = .46$ ).	No
Brannan (2009)	Body Parts Satisfaction Scale – Revised (Body Subscale)	F-MPS	- CoM - PS - PE - PC	Correlation.	CoM was associated with body satisfaction ( $r = -.381$ ).	No.
Brosof (2017)	Social Appearance Anxiety Scale	F-MPS	- CoM	Correlation and mediation.	SAA was associated with CoM at all three time points ( $r_s = .33$ to $.52$ ).  <u>Mediation:</u> T1 CoM accounted for a significant amount of variance in T2 SAA ( $\beta = .11$ ). T2 CoM accounted for a significant amount of variance in T3 SAA ( $\beta = .20$ ).	<u>Mediation:</u> Dietary restraint, binge eating, and previous T1/2 scores.
Buhlmann (2008)	BDD Modification of the Yale-Brown OCD Scale	F-MPS	- O (not included in F-MPS total score) - CoM - DaA - PC - PE - PS	One-way ANOVA.	BDD group > control group on scores for F-MPS ( $d = 0.86$ ), CoM ( $d = 1.20$ ), and DaA ( $d = 1.58$ ).  Only difference between clinical groups was on DaA score - OCD group > BDD group ( $d = \text{NR}$ ).	No
Cash (1995)	Body Image Ideals Questionnaire	F-MPS	None.	Correlation and partial correlation.	Negative body image was associated with perfectionism ( $r = .28$ ).  <u>Partial Correlation:</u> This remained significant ( $r = .25$ ).	<u>Partial correlation:</u> Social desirability.

1 <sup>st</sup> Author and Year	Measure of Appearance-Related Concerns	Measure of Perfectionism	Dimensions of Perfectionism	Statistical Analysis	Significant Findings ( $p < .05$ )*	Other Covariates
Dakanalis (2014)	Male Body Dissatisfaction Scale  Muscle Dysmorphia Disorder Inventory (Appearance Intolerance subscale)	F-MPS	- CoM - PS - DaA	Partial Correlation.	Body dissatisfaction was associated with CoM ( $r = .54$ ), PS ( $r = .55$ ), and DaA ( $r = .56$ ).  Appearance intolerance was associated with CoM ( $r = .55$ ), PS ( $r = .60$ ), and DaA ( $r = .49$ ).	<u>Partial correlation:</u> BMI
Frederick (2016)	MBRSQ (Appearance Evaluation Subscale)	F-MPS	- CoM - PS - PE - PC	Correlation, regression.	Asian American women: Positive appearance evaluation was associated with PE ( $r = -.16$ ), PC ( $r = -.30$ ), and CoM ( $r = -.36$ ).  White American women: Positive appearance evaluation was associated with PE ( $r = -.13$ ), PC ( $r = -.26$ ), and CoM ( $r = -.31$ ).  <u>Regression:</u> PS and CoM accounted for a significant amount of variance in appearance evaluation ( $\beta = .13$ and $\beta = -.17$ respectively). This lost significance when face image measures were included in the model.	<u>Regression:</u> Ethnicity, BMI, age, body surveillance, sociocultural influences, and construal of self (plus face satisfaction in an additional model).
Hartmann (2014)	BDD Modification of the Yale-Brown OCD Scale	F-MPS	- O - CoM - DaA - PC - PE - PS	ANOVA and ANCOVA.	<u>ANOVA:</u> BDD group > control group on scores for all scales of F-MPS (except O) ( $0.007 \geq \omega^2 \geq 0.074$ ). AN group > BDD group on DaA score.	<u>ANCOVA:</u> Depression

1 <sup>st</sup> Author and Year	Measure of Appearance-Related Concerns	Measure of Perfectionism	Dimensions of Perfectionism	Statistical Analysis	Significant Findings ( $p < .05$ )*	Other Covariates
					ANCOVA: BDD group > control group on perfectionism score ( $\omega^2 = \text{NR}$ ).	
Lamanna (2010)	Male Body Attitudes Scale  Muscle Dysmorphia Disorder Inventory	F-MPS	- O - CoM - DaA - PC - PE - PS	Correlation and path analysis.	Men: Overall perfectionism was associated with muscle dysmorphia ( $r = .413$ ) and body dissatisfaction ( $r = .412$ ).  Women: Overall perfectionism was associated with muscle dysmorphia (in a confirmatory sample only: $r = .415$ ).  <u>Path Analysis:</u> Four F-MPS subscales accounted for a significant amount of variance in male body dissatisfaction: CoM ( $\beta = .396$ ), DaA ( $\beta = .403$ ), PC ( $\beta = .419$ ), and PE ( $\beta = .342$ ).	<u>Path analysis:</u> Various pathways between sociocultural influence, self-esteem, body dissatisfaction, negative affect, muscle dysmorphia.
Levinson (2016)	Social Appearance Anxiety Scale	F-MPS	- Maladaptive - High Standards	Correlation and equation modelling.	T1 maladaptive perfectionism was associated with T1 and T2 SAA (both $r = .37$ ).  T2 maladaptive perfectionism was associated with T1 and T2 SAA ( $r = .33$ and $r = .47$ respectively).  <u>Prospective Model Analysis:</u> Maladaptive perfectionism had a direct effect on T2 SAA ( $\beta = .24$ ).	<u>Prospective Model Analysis:</u> ED symptoms, social anxiety, high standards, fear of negative evaluation, T1 scores.
Levinson (2013) STUDY 1	Social Appearance Anxiety Scale	F-MPS	- Maladaptive - High Standards	Correlation.	Maladaptive perfectionism was associated with SAA ( $r = .35$ ).	No.

1 <sup>st</sup> Author and Year	Measure of Appearance-Related Concerns	Measure of Perfectionism	Dimensions of Perfectionism	Statistical Analysis	Significant Findings ( $p < .05$ )*	Other Covariates
STUDY 2	Social Appearance Anxiety Scale	F-MPS	- Maladaptive - High Standards	Correlation.	Maladaptive perfectionism was associated with SAA ( $r = .37$ ).	No.
Malkina-Pykh (2012)	Body Image Test	F-MPS	- CoM/DaA - PS	Correlation (spearman).	Body image dissatisfaction was associated with PS ( $r = .41$ ) and CoM/DaA ( $r = .45$ ).	No.
Menatti (2013)	Assessment of Body-Image Cognitive Distortions	F-MPS	- Maladaptive - High Standards	Correlation.	Maladaptive perfectionism was associated with Maladaptive body-image cognitions ( $r = .41$ ).	No.

*Note:* F-MPS = Frost Multidimensional Perfectionism Scale; O = Organisation subscale; CoM = Concerns over Mistakes subscale; DaA – Doubts about Actions subscale; PS = Personal Standards subscale; PE = Parental Expectations subscale; PC = Parental Criticism subscale; PP = Parental Perceptions subscale (PE and PC); Maladaptive = CoM, DaA, PC, and PE subscales; High Standards = PS subscale; MBSRQ = Multidimensional Body-Self Relations Questionnaire; ANOVA = Analysis of Variance; ANCOVA = Analysis of Covariance; AN = Anorexia; ED = Eating Disorder; SAA = Social Appearance Anxiety; T1 = Time one; T2 = Time two; BMI = Body Mass Index; NR = Not Reported.

\* Effect sizes are reported in parentheses where available.

Mixed evidence exists for the Personal Standards (PS) dimension, as seven studies included this subscale, yet only two found significant correlations with appearance-related concerns. Despite not finding a correlation between PS and appearance evaluation, Frederick and colleagues did observe that PS accounted for a significant amount of the variance (Frederick, Kelly, Latner & Sandhu, 2016). Similarly, four studies split the dimensions into the “maladaptive” and “high standards” categories, where all four reported no significant associations between high standards (i.e. PS) and appearance-related concerns, but did find maladaptive perfectionism to be associated (i.e. CoM, DaA, Parental Concern [PC], and Parental Expectations [PE]). Further, both the PE and PC dimensions obtained inconsistent findings, with half of the four studies observing significant results.

The final of the F-MPS subscales, Organisation (O), was included independently within three studies, although it was consistently observed to have no relationship with appearance-related concerns.

#### HF-MPS

Fourteen studies measured perfectionism using the HF-MPS, and the findings are presented in Table 5 below. Two of these studies only used one of the measure’s dimensions, whilst six compared all three, and the other six compared two (typically Self-Oriented (SOP) and Socially-Prescribed (SPP) perfectionism, one compared SPP and OOP). All studies found significant observations between at least one of the HF-MPS dimensions and appearance-related concerns.

The SPP dimension of perfectionism was the scale most consistently found to yield stronger associations with appearance-related concerns compared to SOP and OOP. Additionally, using an analogue sample, a “BDD” group were found to have significantly higher scores on SPP (and not SOP) compared to controls. Two studies noted SOP to have a stronger correlation compared to SPP and OOP, which both used the appearance orientation subscale of the MBSRQ (Barnes & Caltabiano, 2017; Dunn, Craft, Dunn & Gotwals, 2011). Otherwise, the association between SOP and other measures of appearance-related concerns fell inferior to SPP ( $N = 7$ ) or did not reach significant at all ( $N = 3$ ). Findings on the association between OOP and appearance-related concerns were mixed. Four of seven studies found that OOP had no significant association with appearance concerns, whereas three reported significant correlations. Although this association was typically comparable or not as strong compared to SPP/SOP.



**Table 5: Summary of the study findings from those using the Hewitt & Flett Multidimension Perfectionism Scale (*N* = 14).**

1 <sup>st</sup> Author and Year	Measure of Appearance-Related Concerns	Measure of Perfectionism	Dimensions of Perfectionism	Statistical Analysis	Significant Findings ( <i>p</i> < .05)*	Other Covariates
Bardone-Cone (2008)	Appearance Self-Esteem	HF-MPS	- SPP	Correlation, regression.	Appearance self-esteem was associated with SPP in men ( <i>r</i> = -.28) and women ( <i>r</i> = -.33).  <u>Regression:</u> SPP accounted for a significant amount of variance in appearance self-esteem in women only ( $\beta$ = .19).	<u>Regression:</u> 4 subscales of the Sociocultural Attitudes, weight-related teasing, and BMI.
Barnes (2017)	MBSRQ (appearance orientation and appearance evaluation subscales)	HF-MPS	- SPP - SOP - OOP	Correlation.	Importance attributed to appearance was associated with SPP ( <i>r</i> = .25); SOP ( <i>r</i> = .48); and OOP ( <i>r</i> = .29).  More favourable evaluation of appearance was associated with SPP ( <i>r</i> = .18).	No
Bartsch (2007)	Body Dysmorphic Disorder Questionnaire  Dysmorphic Concerns Questionnaire	HF-MPS	- SPP - SOP	Correlation, regression, and Mann-Whitney U-Tests.	Dysmorphic concern was associated with SOP ( <i>r</i> = .22) and SPP ( <i>r</i> = .36).  <u>Regression:</u> SOP and SPP accounted for a significant amount of variance in dysmorphic concern ( $\beta$ = .09 and $\beta$ = .10 respectively).  <u>Mann-Whitney U-Test:</u> “Probable BDD” group > non-BDD group on SPP score.	<u>Regression:</u> Self-esteem, depression, and gender.
Brannan (2008)	Body Parts Satisfaction Scale – Revised (Body Subscale)	HF-MPS	- SPP - SOP	Correlation.	Body satisfaction was associated with SPP only ( <i>r</i> = -.25).	No

1 <sup>st</sup> Author and Year	Measure of Appearance-Related Concerns	Measure of Perfectionism	Dimensions of Perfectionism	Statistical Analysis	Significant Findings ( $p < .05$ )*	Other Covariates
Davis (1997)	Body Esteem Scale	HF-MPS	- SPP - SOP - OOP	Correlation, regression.	Body esteem was associated with SOP ( $r = -.46$ ) and SPP ( $r = -.49$ ).  <u>Regression:</u> SOP accounted for a significant amount of variance in body esteem ( $\beta = \text{NR}$ ).	<u>Regression:</u> Neuroticism.
Davis (2005)	Drive for Muscularity Scale  MBSRQ (Appearance Orientation Subscale)	HF-MPS	- SOP	Correlation, regression.	Drive for muscularity and appearance-related attitudes/behaviours were associated with SOP ( $r = .31$ and $r = .23$ respectively).  <u>Regression:</u> SOP accounted for a significant amount of variance in drive for muscularity ( $\beta = .19$ ).	<u>Regression:</u> Neuroticism, appearance orientation, fitness orientation, and BMI.
Dryer (2016)	Muscle Dysmorphia Questionnaire	HF-MPS	- SPP - SOP	Correlation (Spearman), mediation.	Muscle dysmorphia was associated with SOP ( $r = .19$ ) and SPP ( $r = .30$ ).  <u>Mediation (standardised coefficients NR):</u> SOP had a direct effect on muscle dysmorphia only when teasing was used as the IV.  SPP had a direct effect on muscle dysmorphia, regardless of whether media influence, peer influence, or teasing were used as the IV.	<u>Mediation:</u> Age, BMI, media influence, peer influence, teasing.
Dunn (2011)	MBSRQ (Appearance orientation, appearance evaluation, and body area satisfaction Subscales)	HF-MPS	- SPP - SOP - OOP	Correlation, regression.	SOP was associated with importance attributed to appearance ( $r = .42$ ); body area satisfaction ( $r = -.21$ ); and negative attitudes towards body image ( $r = .32$ ).  SPP was associated with adaptive evaluation of appearance ( $r = -.30$ ), importance attributed to appearance ( $r = .33$ ), body area	No.

1 <sup>st</sup> Author and Year	Measure of Appearance-Related Concerns	Measure of Perfectionism	Dimensions of Perfectionism	Statistical Analysis	Significant Findings ( $p < .05$ )*	Other Covariates
	Body Image Ideal Questionnaire				satisfaction ( $r = -.35$ ), and negative attitudes towards body image ( $r = .35$ ).  <u>Regression:</u> Total perfectionism accounted for significant variance in importance attributed to appearance, evaluation of appearance, body area satisfaction, and negative attitudes towards body image ( $\beta = \text{NR}$ ).	
Ferrand (2007)	Body Esteem Scale (Appearance and Attribution subscale)	HF-MPS	- SPP - SOP	Correlation.	Body esteem was associated with SPP ( $r = -.42$ ).	No.
Grammas (2009)	Male Body Attitudes Scale (Muscularity subscale)	HF-MPS	- SPP - SOP - OOP	Correlation, regression.	Dissatisfaction in muscularity correlated with SOP ( $r = .18$ ) and SPP ( $r = .21$ ).  <u>Regression:</u> SPP accounted for a significant amount of variance in muscularity ( $\beta = .21$ ).	<u>Regression:</u> Ethnicity, sociocultural influences.
Hanstock (2002)	Dysmorphic Concerns Questionnaire	HF-MPS	- SPP - SOP - OOP	Correlation, regression.	Dysmorphic concern was associated with SPP ( $r = .75$ ).  <u>Regression:</u> SPP accounted for a significant amount of variance in dysmorphic concern ( $\beta = .45$ ).	<u>Regression:</u> GHQ, past acne, GHQ*past acne interaction, and acne quality of life.
Malkina-Pykh (2012)	Body Image Test	HF-MPS	- SPP - OOP	Correlation (Spearman).	Body image dissatisfaction was associated with OOP ( $r = .38$ ) and SPP ( $r = .43$ ).	No.
Murray (2013)	Muscle Dysmorphia Disorder Inventory	HF-MPS	- SPP - SOP - OOP	Correlation, regression.	Muscle dysmorphia was associated with SOP ( $r = .47$ ), SPP ( $r = .59$ ), and OOP ( $r = .36$ ).	<u>Regression:</u> Interpersonal problems,

1 <sup>st</sup> Author and Year		Measure of Appearance-Related Concerns	Measure of Perfectionism	Dimensions of Perfectionism	Statistical Analysis	Significant Findings ( $p < .05$ )*	Other Covariates
						<p><u>Regression:</u> SOP and SPP accounted for a significant amount of variance in muscle dysmorphia (both <math>\beta = .18</math>).</p> <p>Only SOP and OOP accounted for a significant amount of variance in the functional impairment subscale (<math>\beta = .27</math> and <math>\beta = .25</math> respectively).</p>	difficulties in emotional regulation, self-esteem.
Sherry (2009)	STUDY 2	Appearance Schemas Inventory	HF-MPS	- SPP - SOP	Correlation, regression.	<p>Distorted appearance schemas are associated with SOP (<math>r = .23</math>) and SPP (<math>r = .33</math>).</p> <p>Body image disturbance is associated with SPP (<math>r = .38</math>).</p> <p><u>Regression:</u> SOP and SPP accounted for a significant amount of variance in body image disturbance (<math>\beta = .07</math> and <math>\beta = .23</math> respectively).</p>	<u>Regression:</u> Gender, BMI, reassurance seeking.

*Note:* HF-MPS = Hewitt & Flett Multidimensional Perfectionism Scale; SOP = Self-Oriented Perfectionism; SPP = Socially-Prescribed Perfectionism; OOP = Other-Oriented Perfectionism; MBSRQ = Multidimensional Body-Self Relations Questionnaire; NR = Not reported; ANOVA = Analysis of Variance; BMI = Body Mass Index; GHQ = General Health Questionnaire.

\* Effect sizes are reported in parentheses where available.

### EDI-P

Four studies rated as “fair” quality did not observe significant associations between the EDI-P and any appearance-related concerns (Table 6). Otherwise, the EDI-P (a unidimensional tool) was found to correlate with drive for muscularity, body checking, dysfunctional muscularity attitudes, and dysmorphic concerns. The EDI-P was also found to have a main effect on drive for muscularity, account for a significant amount of variance in dysmorphic concern; and a “BDD” group from an analogue sample scored higher on this measure compared to controls.

### APS-R

The APS-R is a multidimensional tool, however three of the four studies that utilised it only included one dimension (the discrepancy/maladaptive subscale). The remaining study used total score only. All studies were rated as having “fair” quality and reported significant associations, although it is not clear how these compare with the other dimensions from this measure, as no comparisons could be made (Table 7).

### Other Measures of Perfectionism

Table 8 highlights the remaining studies that used a variety of measures of perfectionism. Four out of nine studies measured perfectionism along various unidimensional tools, and therefore cannot be directly compared to each other. Nonetheless they all yielded significant findings. Two studies rated as “fair” quality utilised the OBQ-PC measure, where both obtained significant correlations with body image disturbance and appearance anxiety, however only one found the OBQ-PC to account for a significant amount of variance in appearance-related concern (Blakey et al., 2016).

Two studies by the same author measured perfectionism using the PSPS and included all three of its subscales. Non-display of imperfections was found to be the greatest predictor of body image disturbance in both samples. This was also shown to have the strongest correlation compared to the other two dimensions with body image disturbance in both studies, and distorted appearance schemas in community fitness members. Self-promotion was observed to be the strongest correlate of distorted appearance schemas in university fitness students. Finally, one study rated as “poor” quality utilised the PANPS and found significant correlations between positive and negative perfectionism and dysmorphic concern, although failed to report the correlation statistic.

**Table 6: Summary of the study findings from those using the Eating Disorders Inventory - Perfectionism Subscale ( $N = 7$ ).**

1 <sup>st</sup> Author and Year	Measure of Appearance-Related Concerns	Measure of Perfectionism	Dimensions of Perfectionism	Statistical Analysis	Significant Findings ( $p < .05$ )*	Other Covariates
Chen (2010)	MBSRQ (Appearance Evaluation subscale)	EDI-P	None.	Correlation.	No.	No.
Cooley (2007)	Body Self-Relations Questionnaire	EDI-P	None.	Correlation.	No.	No.
Custers (2009)	Physical Self-Description Questionnaire (Appearance Subscale)	EDI-P	None.	Correlation and partial correlation.	No.	<u>Partial correlation:</u> BMI, school grade, attitudes towards “pro-ana” websites.
Dakanalis (2015 <sub>a</sub> )	Yelland & Tiggerman’s Drive for Muscularity Scale  Male Body Checking Questionnaire	EDI-P	None.	Correlation and moderation analysis.	EDI-P was associated with drive for muscularity ( $r = .47$ ) and body checking ( $r = .20$ ).  <u>Moderation:</u> EDI-P had a main effect on drive for muscularity ( $\beta = .43$ ).	<u>Moderation:</u> Body dissatisfaction, attachment, body checking, emotional dysregulation, and self-esteem.
Minnich (2014)	Drive for Muscularity Scale (Attitudes Subscale)	EDI-P	None.	Correlation.	EDI-P was associated with dysfunctional muscularity attitudes ( $r = .35$ ).	No.
Schieber (2013)	Dysmorphic Concerns Questionnaire	EDI-P	None.	Correlation, regression and ANCOVA.	Dysmorphic concerns were associated with EDI-P ( $r = .24$ ).	<u>ANCOVA:</u> Gender and age.

1 <sup>st</sup> Author and Year	Measure of Appearance-Related Concerns	Measure of Perfectionism	Dimensions of Perfectionism	Statistical Analysis	Significant Findings ( $p < .05$ )*	Other Covariates
					<u>Regression:</u> Perfectionism accounted for a significant amount of variance in dysmorphic concerns ( $\beta = .15$ ).  <u>ANCOVA:</u> BDD group > control group on EDI-P score ( $d = 0.43$ ).	<u>Regression:</u> Aesthetic sensitivity and behavioural inhibition.
Shaw (2004)	Body Parts Satisfaction Scale	EDI-P	None.	Correlation.	No.	No.

*Note:* EDI-P = Eating Disorders Inventory – Perfectionism Subscale; MBSRQ = Multidimensional Body-Self Relations Questionnaire; ANCOVA = Analysis of Covariance; BMI = Body Mass Index.

\* Effect sizes are reported in parentheses where available.

**Table 7: Summary of the study findings from those using the Almost Perfect Scale - Revised ( $N = 4$ ).**

1 <sup>st</sup> Author and Year		Measure of Appearance-Related Concerns	Measure of Perfectionism	Dimensions of Perfectionism	Statistical Analysis	Significant Findings ( $p < .05$ )*	Other Covariates
Barnett (2016)	STUDY 1	Body Image Satisfaction Scale (Holsen Version)	APS-R	- Maladaptive (Discrepancy)	Correlation and mediation.	Body satisfaction was associated with maladaptive perfectionism ( $r = -.39$ ).  <u>Mediation:</u> Maladaptive perfectionism had a negative direct effect on body image satisfaction in both a single and multiple model (CD if coefficients are standardised).	<u>Mediation:</u> (Single) Self-compassion. (Multiple) 6 subscales of the Self-Compassion Scale.
	STUDY 2	Body Image Satisfaction Scale (Holsen Version)	APS-R	- Maladaptive (Discrepancy)	Correlation and mediation.	Body satisfaction was associated with maladaptive perfectionism ( $r = -.28$ ).  <u>Mediation:</u> Maladaptive perfectionism had a negative direct effect on body image satisfaction in both a single and multiple model (CD if coefficients are standardised).	<u>Mediation:</u> (Single) Self-compassion. (Multiple) 6 subscales of the Self-Compassion Scale.
	Dour (2011)	Body Esteem Scale for Adolescents and Adults (Appearance and Attribution Subscales)	APS-R	- Maladaptive (Discrepancy)	Correlation and regression.	Appearance esteem was associated with maladaptive perfectionism ( $r = CD$ ).  <u>Regression:</u> Maladaptive perfectionism accounted for a significant amount of variance in appearance esteem in females ( $\beta = -.43$ ) and males ( $\beta = -.42$ ).  Maladaptive perfectionism accounted for a significant amount of variance in perceived evaluation by others in females ( $\beta = -.28$ ) and males ( $\beta = -.32$ ).	<u>Regression:</u> Endorsement of superhero ideal.



1 <sup>st</sup> Author and Year	Measure of Appearance-Related Concerns	Measure of Perfectionism	Dimensions of Perfectionism	Statistical Analysis	Significant Findings ( $p < .05$ )*	Other Covariates
Midlarsky (2008)	Aging-Related Concerns about Appearance	APS-R	None.	Correlation.	Perfectionism was associated with aging-related appearance concerns ( $r = .43$ ).	No.

*Note:* APS-R = Almost Perfect Scale - Revised; CD = Cannot Determine.

\* Effect sizes are reported in parentheses where available.

**Table 8: Summary of the study findings from those using other measures of perfectionism ( $N = 9$ ).**

1 <sup>st</sup> Author and Year		Measure of Appearance-Related Concerns	Measure of Perfectionism	Dimensions of Perfectionism	Statistical Analysis	Significant Findings ( $p < .05$ )*	Other Covariates
Rudiger (2007)		Body Image States Scale	PSPS	None.	Correlation and regression.	Perfectionism was associated with day-to-day variability in body image ( $r = .201$ ).	<u>Regression:</u> Appearance schemas and appearance cognitive distortions.
Sherry (2009)	STUDY 1	Body Image Rating Scale  Appearance Schemas Inventory	PSPS	- Perfectionistic Self-Promotion (SP) - Non-disclosure of Imperfection (NDisc) - Non-display of Imperfection (NDisp)	Correlation, regression, and mediation	Distorted appearance schemas were associated with SP ( $r = .63$ ), NDisc ( $r = .51$ ), and NDisp ( $r = .65$ ). Body image disturbance was associated with SP ( $r = .51$ ), NDisc ( $r = .48$ ), and NDisp ( $r = .58$ ).  <u>Regression:</u> SP ( $\beta = .14$ ), NDisc ( $\beta = .26$ ), and NDisp ( $\beta = .22$ ) accounted for a significant amount of variance in body image disturbance.  <u>Mediation:</u> NDisc was directly linked to body image disturbance ( $\beta = .56$ ) and distorted appearance schemas ( $\beta = .65$ ).	<u>Regression:</u> Gender, BMI, reassurance seeking.
	STUDY 2	Body Image Rating Scale  Appearance Schemas Inventory	PSPS	- SP - NDisc - NDisp	Correlation, regression, and mediation	Distorted appearance schemas were associated with SP ( $r = .58$ ), NDisc ( $r = .40$ ), and NDisp ( $r = .53$ ). Body image disturbance was associated with SP ( $r = .45$ ), NDisc ( $r = .34$ ), and NDisp ( $r = .52$ ).	<u>Regression:</u> Gender, BMI, reassurance seeking.

1 <sup>st</sup> Author and Year	Measure of Appearance-Related Concerns	Measure of Perfectionism	Dimensions of Perfectionism	Statistical Analysis	Significant Findings ( $p < .05$ )*	Other Covariates
					<u>Regression:</u> SP ( $\beta = .04$ ), NDisc ( $\beta = .02$ ), and NDisp ( $\beta = .44$ ) accounted for a significant amount of variance in body image disturbance.  <u>Mediation:</u> NDisp was directly linked to body image disturbance ( $\beta = .53$ ), and distorted appearance schemas ( $\beta = .52$ ).	
Blakey (2016)	Body Image Disturbance Questionnaire	OBQ-PC	None.	Correlation and regression.	Body image disturbance was associated with perfectionism/need for certainty ( $r = .28$ ).  <u>Regression:</u> OBQ-PC accounted for a significant amount of variance in body image disturbance ( $\beta = .13$ ).	<u>Regression:</u> Gender, obsessive-compulsive scale, depression, eating attitudes, other OBQ subscales.
Lavell (2014)	Appearance Anxiety Inventory	OBQ-PC	None.	Correlation and regression.	Perfectionism/need for certainty was associated with BDD symptoms ( $r = .37$ ).	<u>Regression:</u> Age, gender, social anxiety, thought control, vigilance, delusional beliefs.
Rasooli (2011)	Body Image Concerns Inventory	PANPS	- Positive - Negative	Correlation and regression.	Dysmorphic concern was positively correlated with negative perfectionism and negatively correlated with positive perfectionism ( $r_s = NR$ ).	<u>Regression:</u> Consciousness, agreeableness.

1 <sup>st</sup> Author and Year	Measure of Appearance-Related Concerns	Measure of Perfectionism	Dimensions of Perfectionism	Statistical Analysis	Significant Findings ( $p < .05$ )*	Other Covariates
					<u>Regression:</u> Negative and positive perfectionism accounted for a proportion of the variance in dysmorphic concern ( $\beta$ s = NR).	
Davis (1997)	Body Esteem Scale	NPQ	None.	Correlation, regression.	Body esteem was associated with neurotic perfectionism ( $r = -.57$ ).  <u>Regression:</u> Neurotic perfectionism accounted for a significant amount of variance in body esteem ( $\beta = \text{NR}$ ).	<u>Regression:</u> Neuroticism.
Frederick (2016)	MBSRQ (Appearance Evaluation Subscale)	Perfectionism Inventory	- Need for approval	Correlation and regression.	Appearance evaluation was associated with need for approval in Asian American ( $r = -.34$ ) and White American ( $r = -.36$ ) women.	<u>Regression:</u> Ethnicity, BMI, age, body surveillance, sociocultural influences, and construal of self. Face satisfaction also in additional model.
Saling (2005)	Body Change Inventory (Muscle Preoccupation Subscale)	CAPS	None.	Regression.	Cross-sectional: Perfectionism accounted for a significant amount of variance in muscle preoccupation in girls ( $\beta = .28$ ) and boys ( $\beta = .36$ ).	<u>Regression:</u> BMI, grade, self-esteem, parental and peer relations, negative affect. Time 1 scores also

1 <sup>st</sup> Author and Year	Measure of Appearance-Related Concerns	Measure of Perfectionism	Dimensions of Perfectionism	Statistical Analysis	Significant Findings ( $p < .05$ )*	Other Covariates
					Longitudinal: Perfectionism accounted for a significant amount of variance in muscle preoccupation in boys only ( $\theta = .20$ ).	for longitudinal analysis.

*Note:* PSPS = Perfectionistic Self-Presentation Scale; OBQ-PC = Obsessive Beliefs Questionnaire – Perfectionism and Need for Certainty Subscale; PANPS = Positive And Negative Perfectionism Scale; NPQ = Neurotic Perfectionism Questionnaire; CAPS = Child and Adolescent Perfectionism Scale; MBSRQ = Multidimensional Body-Self Relations Questionnaire; NR = Not Reported; BMI = Body Mass Index.

\* Effect sizes are reported in parentheses where available.

### 3.5.3. Does the association still exist when controlling for other mental health diagnoses?

Overall, 31 of the studies (70%) conducted analysis that controlled for at least one other variable. Nine studies (20%) controlled for variables associated with mental health diagnoses, including negative affect/low mood/depression, symptoms of eating disorders, muscle dysmorphia, social anxiety, OCD, or general mental health. The most common mental health symptom controlled for was depression (or equivalent), which featured in five of the studies. Only one study controlled for symptoms of three separate mental health diagnoses simultaneously (OCD, depression, and disordered eating), and six studies controlled for only one.

Of the nine studies that controlled for at least one other psychopathology, six found that their significant results remained after accounting for co-occurring psychopathology. Two studies rated as “fair” quality found varying results after controlling for mental health diagnoses. For instance, Hartmann and colleagues (Hartmann, Thomas, Greenberg, Matheny & Wilhelm, 2014) found that a BDD patient group scored significantly higher than a control group on various dimensions of perfectionism, which remained significant after controlling for depression, although an initial significant difference between the BDD group and Anorexia group in DaA scores did not maintain significance after including depression into the model. Similarly, another study (Lamanna, Grieve, Derryberry, Hakman & McClure, 2010) found total perfectionism to be correlated with body dissatisfaction in men. However, total perfectionism was not found to significantly predict body dissatisfaction (whilst controlling for other pathways that included negative affect and muscle dysmorphia), yet in their discussion the researchers comment that four of the subscales did (CoM, DaA, PC, and PE). One study (rated as “fair” quality) also failed to observe significant regression results (where the model included social anxiety), despite detecting a significant correlation between perfectionism and symptoms of BDD initially (Lavell, Farrell, & Zimmer-Gembeck, 2014).

### 3.5.4. Can any causal conclusions be drawn?

Only three studies were designed in a manner to address this research question. Perfectionism was observed to prospectively and significantly predict social appearance anxiety between two to six months after baseline (Brosof, Cheri & Levinson, 2017; Levinson & Rodebaugh, 2016), and accounted for a significant amount of variance in muscle preoccupation in boys (not girls) after a six-month period (Saling et al., 2005).

#### 3.5.5. Do the associations vary depending on primary diagnosis?

Three papers recruited from psychiatric samples (Buhlmann, Etcoff, & Wilhelm, 2008; Davis, 1997; Hartmann et al., 2014), however only two of these included a BDD and control comparison group (Buhlmann et al., 2008; Hartmann et al., 2014). Of the available information, it appears that clinical groups (BDD, OCD, and Anorexia) do self-report more global perfectionistic traits compared to controls. Buhlmann and colleagues (2008) found that this was also specifically in the CoM and DaA dimensions only, whereas Hartmann and Colleagues (2014) found this in all F-MPS subscales (except O). In both studies (rated as “fair” quality), the BDD group scored significantly lower compared to the OCD/Anorexia group on the F-MPS’s DaA dimension. However, when depression scores were controlled, this difference diminished (Hartmann et al., 2014).

## 4. DISCUSSION

---

The aim of the current review was to consolidate the published literature that has explored the relationship between perfectionism and BDD, which was the first systematic review of its kind. Given the lack of literature available in this area, BDD-associated constructs were also included, such as measures of appearance-related concerns (excluding those that assess weight/shape concerns) and all quantitative methodologies were accepted. In total, 709 studies were screened where 41 published articles met the inclusion criteria (utilising 44 independent samples). The outcome measures adopted to assess perfectionism and BDD/appearance-related concerns were diverse, and so this review also attempted to summarise the information in the most efficient and comprehensible manner, in line with the review's research questions.

### 4.1. Summary

There were several trends observed in the types of studies that have been conducted in this research area to date. Studies were predominantly cross-sectional in design, which meant that their quality was somewhat limited in comparison to other methodological designs (e.g. controlled interventions, case-controls etc.), as the findings are more likely to be susceptible to bias. Nonetheless, these studies typically had good sample sizes although they tended to be enlisted from countries considered to represent “Western” society. Samples were largely recruited from adult, unselected, university populations, although five focused uniquely on child populations. As predicted, there were very few studies explicitly examining the perfectionism-BDD association, with only two studies including patients with formal BDD diagnoses. Almost half the samples exclusively consisted of female participants, with only eight studies recruiting a relatively heterogenous group. Surprisingly, despite ethnicity being a factor known to effect appearance-related concerns (Bartsch, 2007; Bohne, Keuthen, Wilhelm, Deckersbach, & Jenike, 2002; Boroughs, Krawczyk & Thompson, 2010; Cash et al., 2004; Marques et al., 2011; Mayville et al., 1999), a third of studies did not report the ethnic composition of their sample. Of those that did, these were rarely diverse or reflective of ethnic minorities. Moreover, the included studies dated from 1995 – 2017, which is reflective of the lag in research described earlier for this area (Jerome, 2017).



#### 4.1.1. Is there an association between perfectionism and BDD?

In total, 32 different measures of appearance-related concerns were used, suggesting that any conclusions drawn were not exclusively specific to BDD. Additionally, 10 different measures of perfectionism were included in the studies. Ninety per-cent of cross-sectional studies reported significant associations between these variables, where the vast majority of these were rated as being “fair” in quality. According to Cohen (1988),  $r$  values between .0 and .3 indicate small effects, .3 to .5 indicate medium/moderate effects, and  $> .5$  indicate large effects. Therefore, associations between perfectionism and appearance-related concerns varied from a small to large effect. Of interest, one factor linking the four non-significant findings together was their use of the EDI-P to assess perfectionism, which only consists of 6-items and is specific to the context of eating disorders. The two variables of interest were further observed to yield significant results in regression, mediation, and moderation analysis. However, the statistics presented for these analyses ( $\beta$  coefficients) cannot be directly compared to correlation coefficients due to them being conditional on the other variables entered into their models. For the studies that had comparative designs, clinical and analogue groups of BDD were found to express significantly higher levels of perfectionism compared to controls, and significant associations were observed over time. Three of the four studies reported effect sizes and showed small to large effects. Taken together, all but four cross-sectional studies reported significant associations between the construct of perfectionism and appearance-related concerns, despite a wide range of measures being used. Patients with BDD were also demonstrated to score significantly higher on measures of perfectionism compared to controls. Therefore, in the context of largely cross-sectional data rated as “fair” in quality, perfectionism does appear to be reliably associated with appearance-related concerns, with higher levels of perfectionism correlating with greater concerns with appearance.

#### 4.1.2. Are some dimensions of perfectionism more strongly or consistently associated with appearance-related concerns?

Five of the 10 perfectionism tools were multidimensional, where two of these were equally as popular throughout the review, which were the original multidimensional tools developed to assess perfectionism in the 1990s (Shafran et al., 2002). When using the F-MPS, the findings illustrated that Concern over Mistakes (CoM), Doubts about Actions (DaA), or the “maladaptive” dimension of perfectionism were consistently associated with appearance-related concerns, whereas this relationship was variable when Personal Standards (PS), Parental Concerns (PC), or Parental Expectations (PE) was used. This is in line with previous research comparing individuals with OCD

and nonclinical controls (Anthony, Purdon, Huta, & Swinson, 1998; Frost & Steketee, 1997; Sassaroli et al., 2008). Organisation (O) was consistently not related to appearance-concerns when considered independently, which has also been observed in various other studies and is why this dimension is often neglected (Kim, Chen, MacCann, Karlov, & Kleitman, 2015). In the context of the HF-MPS, Socially-Prescribed Perfectionism (SPP) appears to have the most consistent and strongest relationship with appearance-related concerns, another dimension frequently considered to represent maladaptive perfectionism (DiBartolo & Rendón, 2012). The role of Other-Oriented Perfectionism (OOP) in appearance concerns was variable as findings were mixed, although when significant, the strength of the association was typically weaker in comparison to Self-Oriented Perfectionism (SOP) and SPP, which is consistent with previous findings (Frost et al., 1993; Stoeber, 2014).

Although comprising of three dimensions, only one from the ASP-R was included in the studies and therefore no dimension-level comparisons can be made. Nonetheless, this “maladaptive” dimension was consistently found to be associated with appearance-related concerns. There was some promising evidence for the PSPS and PANPS, although very few studies utilised these and so the consistency of these findings is currently weak. Taken together, the dimensions of perfectionism that reliably yielded the strongest relationships with appearance-related concerns include the CoM and DaA dimensions of the F-MPS, the SPP dimension of the HF-MPS, and when perfectionism is categorised “maladaptive”. This is not surprising given that these dimensions are frequently found to be associated with other mental health diagnoses, and appearance-related concerns are also in the context of self-criticism and evaluation, rather than being achievement striving, healthy or adaptive. More specifically, each of these dimensions may map on to the specific processes/behaviours observed in patients diagnosed with BDD. For instance, DaA (doubting your own ability to accomplish tasks) could reflect the repetitive behaviours; CoM (being self-critical, interpreting mistakes to mean failure) could contribute to cognitive biases, selective attention and enhanced aesthetic perceptual sensitivity; and SPP (belief that others expect perfection and will be critical if this is not achieved) could account for the unrealistic beliefs about appearance and avoidance of social situations due to fear of negative evaluation.

#### 4.1.3. Does the association still exist when controlling for other mental health diagnoses?

One fifth of studies included other mental health diagnoses as covariates in the statistical analysis, and this was most commonly depression. Six of these studies maintained significant results after controlling for co-occurring psychopathology, and two (both of “fair” quality) observed mixed results. The study that failed to find a significant association after controlling for covariates assessed perfectionism using one subscale of the OBQ ( $N = 246$  University participants) (Lavell et al, 2014). The regression model included: age, gender, social anxiety, thought control, vigilance, delusional beliefs. In contrast, Blakey and colleagues (2016) also used the OBQ but found a significant effect whilst controlling for gender, the other subscales in the OBQ, symptoms of OCD, depression, and eating disorders ( $N = 601$  University participants). A noticeable difference between these studies is the sample size and the covariates used. In sum, of the few studies that did assess and control for covariates associated with mental health diagnoses, the findings suggest that the association between perfectionism and appearance-related concerns does typically exist beyond the influence of these disorders, although more studies that explore and control for this are required.

#### 4.1.4. Can any causal conclusions be drawn?

Three of the 44 studies were designed longitudinally, where positively, measures of perfectionism did prospectively and significantly predict appearance-related concerns over time. However, extensive conclusions could not be drawn on the causality of the relationship, and the follow-up time periods were relatively brief (i.e. no longer than 10 months), although these initial findings warrant further investigations in this area.

#### 4.1.5. Do the associations vary depending on primary diagnosis?

Due to most of the papers being cross-sectional and based on unselected student samples, this question could not be answered sufficiently. In the two studies that did compare different clinical groups, a BDD group did score significantly lower compared to the OCD/Anorexia group on the F-MPS's DaA dimension. However, when depression scores were controlled, this difference diminished (Hartmann et al., 2014).

## 4.2. General Limitations

The current review needs to be interpreted alongside its limitations. Firstly, most studies had cross-sectional designs and were based upon self-report surveys in unselected student samples. Although this is an effective method to gather vast amounts of information with little burden, it is engrained with several biases including those that effect participant response, measurement of variables, and the interpretation (Deming, 1944). This design also lacks control over unmeasured, confounding variables and does not allow for an understanding of the cause and effect. Although in the current review quality ratings were mostly determined to be “fair”, these were rated in the context of only cross-sectional and longitudinal designs, and therefore the quality would reduce should it be compared to other research methodologies. Generalisability of the findings are further limited to the sampled participants and time of the survey. This is particularly problematic given that the initial aim of this review was to understand the association between perfectionism and BDD or its symptoms. Only two studies included formally diagnosed patients with BDD, and so it remains unclear if/how these findings from normative populations generalise to BDD as a clinical disorder.

Additionally, the high variability in the tools used by the extracted papers meant that interpretation and comparison across studies was extremely difficult. The validity of the tools is also debatable. For instance, Shafran and Mansell (2001) questioned whether assessments of perfectionism truly reflect the “classical concept” of perfectionism. They suggest that SOP, PS, and CoM are the closest dimensions to this construct, whereas other dimensions are *relevant* to perfectionism but not integral (e.g. the SPP dimension, as this subscale rates beliefs about other people’s perfectionism and not self-perception of perfectionism). They comment that *“it is undesirable for the construct of perfectionism to be determined by its measures; rather, the concept of perfectionism should be clearly defined and instruments devised to measure it”* (p. 887). Moreover, a main difficulty throughout the search process was the overlap between BDD and EDs constructs. The role of perfectionism is well-establish in the context of EDs, and so it was important that the current review did not replicate these same findings. Both share an altered body image, where if the preoccupation is predominantly focused on being “too fat” or overweight, it does not reflect the BDD diagnosis (Veale & Neziroglu, 2010). Therefore, it was vital that body image measures were excluded should they predominantly include items relating to weight and shape. However, this was not a black-and-white process, as various tools incorporated both these constructs, and the appropriate measures were often extracted from research and models primarily focusing on EDs. As a result, these findings are likely to be somewhat influenced by weight/shape or ED concerns and not uniquely BDD-related.

Finally, there were certain limitations inherent in the methodology of this systematic review. Studies were required to meet specific inclusion criteria. Unpublished studies were excluded, which on the one hand ensures inclusion of more methodologically rigorous studies, but it may also mean that our findings were vulnerable to publication bias, which may have artificially exaggerated positive findings and influenced our conclusion that there is an association between perfectionism and appearance-related concerns. Similarly, approximately 5% of the 709 screened studies were excluded due to their full-text not being published in English. Almost half of these articles were Korean (45%) and articles published in Spanish (18%) and French (11%) also made up a considerable portion of this inaccessible literature (the remaining languages included German, Russian, Italian, Polish, and Croatian). This suggests the presence of a cultural bias, as there is other literature that could contribute to the understanding of this area, however could not be included due to language barriers. Several studies were also excluded despite measuring both perfectionism and appearance-related concerns, because they did not report the specific association between these variables. This reporting bias is understandable when this relationship is not of primary interest, but subsequently means crucial data is being missed. The study selection and data synthesis were also only performed by one researcher, and so due to human error or researcher bias, some eligible studies or key findings may have been missed. Moreover, specific search terms were identified to obtain relevant papers, however, it is possible for these were not comprehensive enough to identify all papers.

### 4.3. Implications & Future Directions

Despite several limitations, this review does raise several insights and considerations that are key in the recommendations for future research and clinical practice. The delay in research relating to BDD comparative to other clinical disorders has already been noted, where cross-sectional research designs have been useful and informative by providing vast detail in a short space of time. These studies typically provide a rationale for theory and psychological models, and to enhance this research area study designs are now required to test these observations. Subsequently it is recommended that, unless exploring a novel concept, future research designs should be more sophisticated to efficiently test the conclusions drawn from cross-sectional papers, for instance using high quality longitudinal and case-control studies. Both perfectionism and appearance-related concerns have also shown to be transdiagnostic, and overlap with depression, EDs, and OCD (Egan et al., 2011; Sassaroli et al., 2008; Veale, 2004; Veale & Neziroglu, 2010), therefore it is

imperative that new research considers this and is inclusive of such clinical groups and healthy controls. Additionally, the diversity of participants require improvement, and so research should steer away from homogeneous university samples, but be inclusive of ethnic minorities, all age groups (especially children), and representative of both genders, as appearance-related concerns have been shown to vary according to these factors (Bohne et al., 2002; Schneider et al., 2016).

Eventually, a meta-analysis in this area would be beneficial to illustrate the collective effect size between perfectionism and BDD/appearance-related concerns across all studies and samples. This was unable to be conducted in the current review due to the high variability in both the appearance-related concern and perfectionism measures. Although possible, this approach would have not provided any clinical meaning. Subsequently, there is a call the development of a “gold standard” tool for assessing BDD/appearance-related concerns, so that can be consistently be used in research. For example, the BDD modification of the Yale-Brown Obsessive-Compulsive Scale (BDD Y-BOCS; Phillips et al., 1997) is considered a robust tool for measuring symptoms and severity of BDD, however this is a semi-structured, clinician-rated measure which might prove difficult to apply to large samples.

Positively, this review has demonstrated the increasing amount of research being conducted to explore BDD (and/or its associated constructs) and its relationship with perfectionism, as has already been done with other mental health disorders. The current findings provide preliminarily support for this association, however the exact role it plays is yet to be understood (e.g. risk/developmental factor or maintenance). It is clear that perfectionism coincides with appearance-related concerns, and therefore clinically it is important that perfectionism is screened for accordingly when assessing clinically significant appearance-related concerns, such as BDD. Psychological formulation and intervention should also consider the role of perfectionism, and it would be clinically meaningful for research to address whether interventions for perfectionism directly impact upon BDD/appearance-related concerns.

#### 4.4. Conclusions

This was the first review attempting to consolidate the findings of research studies that explore the association between perfectionism and BDD (and its associated constructs). The review demonstrated that Concern over Mistakes (CoM), Doubts about Actions (DaA), Socially-Prescribed Perfectionism (SPP), and maladaptive dimensions of perfectionism were the most consistently related to appearance-related outcomes, and that promising evidence had begun to emerge

regarding the maintenance of this association after controlling for various psychopathology. More good quality research is required to make conclusions about causal role of perfectionism and whether differences exist between clinical groups. However, there was a large variety of assessment tools used but little diversity in study participants. Therefore, these findings need to be considered alongside their limitations (namely the lack of generalisability to BDD clinical groups), and there is a call for an improvement in the design and quality of research in this field. Nonetheless, initial steps to understanding this relationship have been made which will benefit clinical practice in this area.

## 5. REFERENCES

---

- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: American Psychiatric Association.
- Antony, M. M., Purdon, C. L., Huta, V., & Swinson, R. P. (1998). Dimensions of perfectionism across the anxiety disorders. *Behaviour Research and Therapy*, 36, 1143–1154.
- Bardone-Cone, A. M., Cass, K. M., & Ford, J. A. (2008). Examining body dissatisfaction in young men within a biopsychosocial framework. *Body Image*, 5, 183-194.
- Bardone-Cone, A. M., Lin, S. L., & Butler, R. M. (2017). Perfectionism and Contingent Self-Worth in Relation to Disordered Eating and Anxiety. *Behavior Therapy*, 48, 380-390.
- Bardone-Cone, A. M., Wonderlich, S. A., Frost, R. O., Bulik, C. M., Mitchell, J. E., Uppala, S., & Simonich, H. (2007) Perfectionism and eating disorders: Current status and future directions. *Clinical Psychology Review*, 27, 384–405.
- Barnes, M. A., & Caltabiano, M. L. (2017). The interrelationship between orthorexia nervosa, perfectionism, body image and attachment style. *Eating and Weight Disorders-Studies on Anorexia Bulimia and Obesity*, 22, 177-184.
- Barnett, M. D., & Sharp, K. J. (2016). Maladaptive perfectionism, body image satisfaction, and disordered eating behaviors among US college women: The mediating role of self-compassion. *Personality and Individual Differences*, 99, 225-234.
- Bartsch, D. (2007). Prevalence of body dysmorphic disorder symptoms and associated clinical features among Australian university students. *Clinical Psychologist*, 11, 16–23.
- Berscheid, E., Walster, E., & Bohrnstedt, G. (1973). The happy American body: A survey report. *Psychology Today*, 7, 119–131.
- Blakey, S. M., Abramowitz, J. S., & Mahaffey, B. L. (2016). Do obsessive beliefs predict body image disturbance? *Journal of obsessive-compulsive and related disorders*, 11, 976-100.
- Bohne, A., Keuthen, N. J., Wilhelm, S., Deckersbach, T., & Jenike, M. A. (2002). Prevalence of Symptoms of Body Dysmorphic Disorder and Its Correlates: A Cross-Cultural Comparison. *Psychosomatics*, 43, 486-490.



- Boroughs, M. S., Krawczyk, R., & Thompson, J. K. (2010). Body Dysmorphic Disorder among Diverse Racial/Ethnic and Sexual Orientation Groups: Prevalence Estimates and Associated Factors. *Sex Roles, 63*, 725–737.
- Brannan, M. E., & Petrie, T. A. (2008). Moderators of the Body Dissatisfaction-Eating Disorder Symptomatology Relationship: Replication and Extension. *Journal of Counselling Psychology, 55*, 263-275.
- Brannan, M., Petrie, T. A., Greenleaf, C., Reel, J., & Carter, J. (2009). The relationship between body dissatisfaction and bulimic symptoms in female collegiate athletes. *Journal of Clinical Sport Psychology, 3*, 103-126.
- Brosof, L. C., & Levinson, C. A. (2017). Social appearance anxiety and dietary restraint as mediators between perfectionism and binge eating: A six month three wave longitudinal study. *Appetite, 108*, 335-342.
- Brown, T. A., Cash, T. F., & Mikulka, P. J. (1990). Attitudinal body-image assessment: Factor analysis of the body-self relations questionnaire. *Journal of Personality Assessment, 55*, 135-144.
- Buhlmann, U., & Hartmann, A. S. (2017). Cognitive and Emotional Processing in Body Dysmorphic Disorder. In K. A. Phillips (Ed.), *Body Dysmorphic Disorder: Advances in Research and Clinical Practice* (pp. 285-298). New York: Oxford University Press.
- Buhlmann, U., Etcoff, N. L., & Wilhelm, S. (2008). Facial attractiveness ratings and perfectionism in body dysmorphic disorder and obsessive-compulsive disorder. *Journal of Anxiety Disorders, 22*, 540-547.
- Cash, T. F. (2007). *Body-image assessments*. Retrieved from <http://www.body-images.com>.
- Cash, T. F., & Labarge, A. S. (1996). Development of the Appearance Schemas Inventory: A new cognitive body-image assessment. *Cognitive Therapy and Research, 20*, 37–50.
- Cash, T. F., & Szymanski, M. L. (1995). The development and validation of the Body-Image Ideals Questionnaire. *Journal of Personality Assessment, 64*, 466-477.
- Cash, T.F., Phillips, K.A., Santos, M.T., & Hrabosky, J.I. (2004). Measuring “negative body image”: Validation of the body image disturbance questionnaire in a non-clinical population. *Body Image: An International Journal of Research, 1*, 363-372.

- Cash, T. F., Fleming, E. C., Alindogan, J., Steadman, L., & Whitehead, A. (2002). Beyond body image as a trait: The development and validation of the Body-Image States Scale. *Eating Disorders: The Journal of Treatment and Prevention*, 10, 103-113.
- Chen, L. J., Fox, K. R., Haase, A. M., & Ku, P. W. (2010). Correlates of body dissatisfaction among Taiwanese adolescents. *Asia Pacific Journal of Clinical Nutrition*, 19, 172-179.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Psychology Press.
- Conroy, M., Menard, W., Fleming-Ives, K., Modha, P., Cerullo, H., & Phillips, K. A. (2008). Prevalence and clinical characteristics of body dysmorphic disorder in an adult inpatient setting. *General Hospital Psychiatry*, 30, 67–72.
- Cooley, E., Toray, T., Valdez, N., & Tee, M. (2007). Risk factors for maladaptive eating patterns in college women. *Eating and Weight Disorders-Studies on Anorexia Bulimia and Obesity*, 12, 132-139.
- Crocker, J., Luhtanen, R. K., Cooper, M. L., & Bouvrette, A. (2003). Contingencies of Self-Worth in College Students: Theory and Measurement. *Journal of Personality and Social Psychology*, 85, 894 –908.
- Custers, K., & Van den Bulck, J. (2009). Viewership of Pro-Anorexia Websites in Seventh, Ninth and Eleventh Graders. *European Eating Disorders Review*, 17, 214-219.
- Dakanalis, A., Favagrossa, L., Clerici, M., Prunas, A., Colmegna, F., Zanetti, M. A., & Riva, G. (2015<sub>a</sub>). Body Dissatisfaction and Eating Disorder Symptomatology: A Latent Structural Equation Modeling Analysis of Moderating Variables in 18-to-28-Year-Old Males. *Journal of Psychology*, 149, 85-112.
- Dakanalis, A., Timko, A., Madeddu, F., Volpato, C., Clerici, M., Riva, G., & Zanetti, A. M. (2015<sub>b</sub>). Are the Male Body Dissatisfaction and Drive for Muscularity Scales reliable and valid instruments? *Journal of Health Psychology*, 20, 48-59.
- Dakanalis, A., Timko, C. A., Clerici, M., Zanetti, M. A., & Riva, G. (2014). Comprehensive examination of the trans-diagnostic cognitive behavioral model of eating disorders in males. *Eating Behaviors*, 15, 63-67.

- Davis, C. (1997). Normal and neurotic perfectionism in eating disorders: An interactive model. *International Journal of Eating Disorders*, 22, 421-426.
- Davis, C., Karvinen, K., & McCreary, D. R. (2005). Personality correlates of a drive for muscularity in young men. *Personality and Individual Differences*, 39, 349-359.
- Deming, W. E. (1944). On Errors in Surveys. *American Sociological Review*, 9, 359-369.
- DiBartolo, P. M., & Rendón, M. J. (2012). A critical examination of the construct of perfectionism and its relationship to mental health in Asian and African Americans using a cross-cultural framework. *Clinical Psychology Review*, 32, 139–152.
- Dour, H. J., & Theran, S. A. (2011). The interaction between the superhero ideal and maladaptive perfectionism as predictors of unhealthy eating attitudes and body esteem. *Body Image*, 8, 93-96.
- Dryer, R., Farr, M., Hiramatsu, I., & Quinton, S. (2016). The Role of Sociocultural Influences on Symptoms of Muscle Dysmorphia and Eating Disorders in Men, and the Mediating Effects of Perfectionism. *Behavioral Medicine*, 42, 174–182.
- Dunn, J. G. H., Craft, J. M., Dunn, J. C., & Gotwals, J. K. (2011). Comparing a domain-specific and global measure of perfectionism in competitive female figure skaters. *Journal of Sport Behavior*, 34, 25-46.
- Egan, S. J., Wade, T. D., & Shafran, R. (2011). Perfectionism as a transdiagnostic process: A clinical review. *Clinical Psychology Review*, 31, 203-212.
- Ferrand, C., Magnan, C., Rouveix, M., & Filaire, E. (2007). Disordered eating, perfectionism and body-esteem of elite synchronized swimmers. *European Journal of Sport Science*, 7, 223-230.
- Feusner, J. D., Neziroglu, F., Wilhelm, S., Mancusi, L., & Bohon, C. (2010). What causes BDD: research findings and a proposed model. *Psychiatric Annals*, 40, 349–355.
- Flett, G. L., Hewitt, P. L., Besser, A., Su, C., Vaillancourt, T., Boucher, D., Munro, Y., Davidson, L. A., & Gale, O. (2016). The Child-Adolescent Perfectionism Scale: Development, psychometric properties, and associations with stress, distress, and psychiatric symptoms. *Journal of Psychoeducational Assessment*, 34, 634-652.
- Franzoi, S. L., & Shields, S. A. (1984). The Body Esteem Scale: Multidimensional structure and sex differences in a college population. *Journal of Personality Assessment*, 48, 173–178.

- Frederick, D. A., Kelly, M. C., Latner, J. D., Sandhu, G., & Tsong, Y. (2016). Body image and face image in Asian American and white women: Examining associations with surveillance, construal of self, perfectionism, and sociocultural pressures. *Body Image, 16*, 113-125.
- Frost, R. O., Marten, P., Lahart, C., & Rosenblate, R. (1990). The dimensions of perfectionism. *Cognitive Therapy and Research, 14*, 449-468.
- Frost, R. O., & Steketee, G. (1997). Perfectionism in obsessive compulsive disorder patients. *Behaviour Research and Therapy, 35*, 291–296.
- Frost, R. O., Heimberg, R. G., Holt, C. S., Mattia, J. I., & Neubauer, A. L. (1993). A comparison of two measures of perfectionism. *Personality and Individual Differences, 14*, 119-126.
- Garner, D. M. (1991). *EDI-2. Eating disorder inventory-2. Professional manual*. Odessa: Psychological Assessment Resources.
- Grammas, D. L., & Schwartz, J. P. (2009). Internalization of messages from society and perfectionism as predictors of male body image. *Body Image, 6*, 31-36.
- Gupta, M. A., & Schork, N. J. (1993). Aging-related concerns and body image: Possible future implications for eating disorders. *International Journal of Eating Disorders, 14*, 481–486.
- Hallsworth, L., Wade, T., & Tiggemann, M. (2005). Individual differences in male body-image: An examination of self-objectification in recreational body builders. *British Journal of Health Psychology, 10*, 453–465.
- Hanstock, T. L. & O'Mahony, J. F. (2002). Perfectionism, acne, and appearance concerns. *Personality and Individual Differences, 32*, 1317-1325.
- Hart, T. A., Flora, D. B., Palyo, S. A., Fresco, D. M., Holle, C., & Heimberg, R. G. (2008). Development and examination of the social appearance anxiety scale. *Assessment, 15*, 48-59.
- Hartmann, A. S., Thomas, J. J., Greenberg, J. L., Matheny, N. L., & Wilhelm, S. (2014). A comparison of self-esteem and perfectionism in anorexia nervosa and body dysmorphic disorder. *Journal of Nervous and Mental Disease, 202*, 883–888.
- Heatherton, T.F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology, 60*, 895-910.

- Hewitt, E. L., & Flett, G. L. (1991). Perfectionism in the self and social contexts: Conceptualization, assessment, and association with psychopathology. *Journal of Personality and Social Psychology, 60*, 456-470.
- Hewitt, P. L., Flett, G. L., Sherry, S. B., Habke, M., Parkin, M., Lam, R. W., McMurtry, B., Ediger, E., Fairlie, P., & Stein, M. B. (2003). The interpersonal expression of perfection: Perfectionistic self-presentation and psychological distress. *Journal of Personality and Social Psychology, 84*, 1303-1325.
- Hildebrandt, T., Langenbucher, J., & Schlundt, D. G. (2004). Muscularity concerns among men: Development of attitudinal and perceptual measures. *Body Image, 1*, 169–181.
- Hildebrandt, T., Walker, C., Alfano, L., Delinsky, S., & Bannon, K. (2010). Development and Validation of a Male Specific Body Checking Questionnaire. *International Journal of Eating Disorders, 43*, 77–87.
- Hill, R. W., Huelsman, T. J., Furr, R. M., Kibler, J., Vicente, B. B., & Kennedy, C. (2004). A new measure of perfectionism: The Perfectionism Inventory. *Journal of Personality Assessment, 82*, 80-91.
- Holsen, I., Jones, D. C., & Birkeland, M. S. (2012). Body image satisfaction among Norwegian adolescents and young adults: A longitudinal study of the influence of interpersonal relationships and BMI. *Body Image, 9*, 201–208.
- Jakatdar, T. A., Cash, T. F., & Engle, E. K. (2006). Body-image thought processes: The development and initial validation of the Assessment of Body-Image Cognitive Distortions. *Body Image, 3*, 325-333.
- Jerome, L. (2017). Body Dysmorphic Disorder over the Past Century. In K. A. Phillips (Ed.), *Body Dysmorphic Disorder: Advances in Research and Clinical Practice* (pp. 19-32). New York: Oxford University Press.
- Kim, L. E., Chen, L., MacCann, C., Karlov, L., & Kleitman, S. (2015). Evidence for three factors of perfectionism: Perfectionistic Strivings, Order, and Perfectionistic Concerns. *Personality and Individual Differences, 84*, 16–22.
- Lamanna, J., Grieve, F. G., Derryberry, W. P., Hakman, M., & McClure, A. (2010). Antecedents of eating disorders and muscle dysmorphia in a non-clinical sample. *Eating and Weight Disorders-Studies on Anorexia Bulimia and Obesity, 15*, E23-E33.

- Lavell, C. H., Farrell, L. J., & Zimmer-Gembeck, M. J. (2014). Do obsessional belief domains relate to body dysmorphic concerns in undergraduate students? *Journal of Obsessive-Compulsive and Related Disorders*, 3, 354-358.
- Levinson, C. A., & Rodebaugh, T. L. (2016). Clarifying the prospective relationships between social anxiety and eating disorder symptoms and underlying vulnerabilities. *Appetite*, 107, 38-46.
- Levinson, C. A., Rodebaugh, T. L., White, E. K., Menatti, A. R., Weeks, J. W., Iacovino, J. M., & Warren, C. S. (2013). Social appearance anxiety, perfectionism, and fear of negative evaluation: distinct or shared risk factors for social anxiety and eating disorders? *Appetite*, 67, 125-133.
- Littleton, H. L., Axsom, D., & Pury, C. L. S. (2005). Development of the body image concern inventory. *Behaviour Research and Therapy*, 43, 229-241.
- Malkina-Pykh, I. G. (2012). Effectiveness of Rhythmic Movement Therapy for Disordered Eating Behaviors and Obesity. *Spanish Journal of Psychology*, 15, 1371-1387.
- Marques, L., LeBlanc, N., Weingarden, H., Greenberg, J. L., Traeger, L. N., Keshaviah, A. & Wilhelm, S. (2011). Body dysmorphic symptoms: Phenomenology and ethnicity. *Body Image*, 8, 163-167.
- Marsh H. W., Richards, G. E., Johnson, S., Roche, L., & Tremayne, P. (1994). Physical Self-Description Questionnaire: Psychometric Properties and a Multitrait-Multimethod Analysis of Relations to Existing Instruments. *Journal of Sport & Exercise Psychology*, 16, 270-305.
- Mayville, S., Katz, R. C., Gipson, M. T., & Cabral, K. (1999). Assessing the prevalence of body dysmorphic disorder in an ethnically diverse group of adolescents. *Journal of Child and Family Studies*, 8, 357-362.
- McCreary, D. R., & Sasse, D. K. (2000). An exploration of the drive for muscularity in adolescent boys and girls. *Journal of American College Health*, 48, 297-304.
- Menatti, A. R., Weeks, J. W., Levinson, C. A., & McGowan, M. M. (2013). Exploring the Relationship Between Social Anxiety and Bulimic Symptoms: Mediation Effects of Perfectionism Among Females. *Cognitive Therapy and Research*, 37, 914-922.
- Mendelson, B. K., Mendelson, M. J., & White, D. R. (2001). Body-esteem scale for adolescents and adults. *Journal of Personality Assessment*, 76, 90-106.

- Midlarsky, E., & Nitzburg, G. (2008). Eating Disorders in Middle-Aged Women. *Journal of General Psychology, 135*, 393-407.
- Minnich, A. M., Gordon, K. H., Holm-Denoma, J. M., & Troop-Gordon, W. (2014). A test of an interactive model of binge eating among undergraduate men. *Eating Behaviors, 15*, 625-631.
- Mitzman, S. F., Slade, P. D., & Dewey, M. E. (1994). Preliminary development of a questionnaire designed to measure neurotic perfectionism in the eating disorders. *Journal of Clinical Psychology, 50*, 516-520.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine, 6*, e1000097. doi:10.1371/journal.pmed1000097.
- Murray, S. B., Rieger, E., Karlov, L., Touyz, S. W. (2013). An Investigation of the Transdiagnostic Model of Eating Disorders in the Context of Muscle Dysmorphia. *European Eating Disorders Review, 21*, 160-164.
- National Institute of Health. (2014). *Study Quality Assessment Tools*. Retrieved October 25, 2017, from <https://www.nhlbi.nih.gov/health-pro/guidelines/in-develop/cardiovascular-risk-reduction/tools/background>.
- Neziroglu, F., Khemlani-Patel, S., & Veale, D. (2008). Social learning theory and cognitive behavioral models of body dysmorphic disorder. *Body Image, 5*, 28–38.
- Obsessive Compulsive Cognitions Working Group. (2005). Psychometric validation of the obsessive belief questionnaire and interpretation of intrusions inventory—Part 2: Factor analyses and testing of a brief version. *Behaviour Research and Therapy, 43*, 1527-1542.
- Oosthuizen, P., Lambert, T., & Castle, D. J. (1998). Dysmorphic concern: Prevalence and associations with clinical variables. *Australian & New Zealand Journal of Psychiatry, 32*, 129–132.
- Petrie, T. A., Tripp, M. M., & Harvey, P. (2002). Factorial and Construct Validity of the Body Parts Satisfaction Scale-Revised: An Examination of Minority and Nonminority Women. *Psychology of Women Quarterly, 26*, 213–221.
- Phillips, K. A. (1991). Body dysmorphic disorder: the distress of imagined ugliness. *American Journal of Psychiatry, 148*, 1138–1149.

- Phillips, K. A. (2005). *The Broken Mirror: Understanding and Treating Body Dysmorphic Disorder (Revised and Expanded Edition)*. New York: Oxford University Press.
- Phillips, K. A. (2017). Classification of Body Dysmorphic Disorder and Relevance for Patient Care. In K. A. Phillips (Ed.), *Body Dysmorphic Disorder: Advances in Research and Clinical Practice* (pp. 33-48). New York: Oxford University Press.
- Phillips, K. A., Hollander, E., Rasmussen, S. A., Aronowitz, B. R., DeCaria, C., & Goodman, W. K. (1997). A severity rating scale for body dysmorphic disorder: development, reliability, and validity of a modified version of the Yale-Brown Obsessive Compulsive Scale. *Psychopharmacology Bulletin*, 33, 17-22.
- Pinto, A., Dargani, N., Wheaton, M. G., Cervoni, C., Rees, C. S., & Egan, S. J. (2017). Perfectionism in obsessive-compulsive disorder and related disorders: What should treating clinicians know? *Journal of Obsessive-Compulsive and Related Disorders*, 12, 102-108.
- Rasmussen, J., Gómez, A. F., & Wilhelm, S. (2017). Cognitive-Behavioral Therapy for Body Dysmorphic Disorder. In K. A. Phillips (Ed.), *Body Dysmorphic Disorder: Advances in Research and Clinical Practice* (pp. 357-378). New York: Oxford University Press.
- Rasooli, S. S., & Lavasani, M. G. (2011). Relationship between personality and perfectionism with body image. *Procedia - Social and Behavioral Sciences*, 15, 1015-1019.
- Ricciardelli, L. A., & McCabe, M. P. (2002). Psychometric evaluation of the Body Change Inventory: an assessment instrument for adolescent boys and girls. *Eating Behaviors*, 3, 45-59.
- Rudiger, J. A., Cash, T. F., Roehrig, M., & Thompson, K. (2007). Day-to-day body-image states: Prospective predictors of intra-individual level and variability. *Body Image*, 4, 1-9.
- Saling, M., Ricciardelli, L. A., & McCabe, M. P. (2005). A prospective study of individual factors in the development of weight and muscle concerns among preadolescent children. *Journal of Youth and Adolescence*, 34, 651-661.
- Santaracchi, E., & Dèttore, D. (2012). Muscle dysmorphia in different degrees of bodybuilding activities: Validation of the Italian version of Muscle Dysmorphia Disorder Inventory and Bodybuilder Image Grid. *Body Image*, 9, 396-403.



- Sassaroli, S., Lauro, L. J. R., Ruggiero, G. M., Mauri, M. C., Vinai, P., & Frost, R. (2008). Perfectionism in depression, obsessive-compulsive disorder and eating disorders. *Behaviour Research and Therapy*, 46, 757–765.
- Schieber, K., Kollei, I., de Zwaan, M., Müller, A., & Martin, A. (2013). Personality traits as vulnerability factors in body dysmorphic disorder. *Psychiatry Research*, 210, 242-246.
- Schneider, S. C., Turner, C. M., Mond, J., & Hudson, J. L. (2016). Prevalence and correlates of body dysmorphic disorder in a community sample of adolescents. *Australian and New Zealand Journal of Psychiatry*, 51, 595-603.
- Shafran, R., & Mansell, W. (2001). Perfectionism and Psychopathology: A Review of Research and Treatment. *Clinical Psychology Review*, 21, 879 – 906.
- Shafran, R., Cooper, Z., & Fairburn, C. G. (2002). Clinical perfectionism: a cognitive–behavioural analysis. *Behaviour Research and Therapy*, 40, 773-791.
- Shaw, H. E., Stice, E., & Springer, D. W. (2004). Perfectionism, body dissatisfaction, and self-esteem in predicting bulimic symptomatology: Lack of replication. *International Journal of Eating Disorders*, 36, 41-47.
- Sherry, S. B., Vriend, J. L., Hewitt, P. L., Sherry, D. L., Flett, G. L., & Wardrop, A. A. (2009). Perfectionism dimensions, appearance schemas, and body image disturbance in community members and university students. *Body Image*, 6, 83-89.
- Slaney, R. B. (2015). *APS Revised*. Retrieved December 01, 2017, from <http://kennethwang.com/aprs/>.
- Slaney, R. B., Rice, K. G., Mobley, M., Trippi, J., & Ashby, J. S. (2001). The revised Almost Perfect Scale. *Measurement and Evaluation in Counseling and Development*, 34, 130–145.
- Stoeber, J. (2014). How other-oriented perfectionism differs from self oriented and socially prescribed perfectionism. *Journal of Psychopathology and Behavioral Assessment*, 36, 329–338.
- Terry-Short, L. A., Owens, R. G., Slade, P. D., & Dewey, M. E. (1995). Positive and negative perfectionism. *Personality and Individual Differences*, 18, 663-668.
- Tylka, T. L., Bergeron, D., & Schwartz, J. P. (2005). Development and psychometric evaluation of the Male Body Attitudes Scale (MBAS). *Body Image*, 2, 161–175.

- Veale, D. (2004). Advances in a cognitive behavioural model of body dysmorphic disorder. *Body Image, 1*, 113-125.
- Veale, D., & Neziroglu, F. (2010). *Body Dysmorphic Disorder: A Treatment Manual*. Retrieved from <https://ebookcentral.proquest.com/lib/kcl/detail.action?docID=485651>.
- Veale, D., Akyüz, E. U., & Hodsoll, J. (2015). Prevalence of body dysmorphic disorder on a psychiatric inpatient ward and the value of a screening question. *Psychiatry Research, 230*, 383-386.
- Veale, D., Eshkevari, E., Kanakam, N., Ellison, N., Costa, A., & Werner, T. (2014). The Appearance Anxiety Inventory: Validation of a Process Measure in the Treatment of Body Dysmorphic Disorder. *Behavioural and Cognitive Psychotherapy, 42*, 605-616.
- Wilhelm, S., & Neziroglu, F. (2002). Cognitive theory of body dysmorphic disorder. In: R. O. Frost & G. Steketee (Eds.), *Cognitive Approaches to Obsessions and Compulsions: Theory, Assessment and Treatment* (pp. 203-214). Amsterdam: Elsevier.
- Wilhelm, S., Phillips, K. A., & Steketee, G. (2013). *A cognitive-behavioral treatment manual for body dysmorphic disorder*. New York: Guilford Press.
- Yelland, C., & Tiggemann, M. (2003). Muscularity and the gay ideal: Body dissatisfaction and disordered eating in homosexual men. *Eating Behaviors, 4*, 107-116.

## 6. APPENDIX

### 6.1. Table summarising the perfectionism assessment measures used by the studies.

Measure	Author(s)	No. of Items	Dimensions	Definition
Hewitt and Flett Multidimensional Perfectionism Scale	Hewitt & Flett, 1991	45	1) Self-Oriented	1) The belief that the individual should be perfect and have high standards and are self-critical if they are not.
			2) Socially Prescribed	2) The belief that others expect the individual to be perfect and perceive others to be critical of them if they are not.
			3) Other Oriented	3) The belief that it is important for others to strive for perfection and are critical towards others if they are not.
Child and Adolescent Perfectionism Scale	Flett et al., 2016	22	1) Self-Oriented	1) As above
			2) Socially Prescribed	2) As above
Frost Multidimensional Perfectionism Scale	Frost et al., 1990	35	1) Personal Standards	1) The standards one places on themselves and their importance for self-evaluation.
			2) Concerns over Mistakes	2) The self-critical, negative reactions to mistakes, and tendency to interpret these as failure.
			3) Organisation	3) The importance placed on order and exactness.
			4) Doubts about Actions	4) Doubt own ability to accomplish tasks and dissatisfaction after evaluation of performance.
			5) Parental Expectations	5) The importance placed on meeting expectations of parents.
			6) Parental Criticism	6) The feeling of being criticised by parents for not measuring up to a standard.

Measure	Author(s)	No. of Items	Dimensions		Definition
Perfectionism Inventory – Need for Approval subscale	Hill et al., 2004	59	1)	Concern Over Mistakes	1) As above.
			2)	High Standards for Others	2) Tendency to hold others to one's own perfectionist ideals.
			3)	Need for Approval	3) Tendency to seek validation from others and to be sensitive to criticism.
			4)	Organisation	4) As above.
			5)	Perceived Parental Pressure	5) Tendency to feel the need to perform perfectly to obtain parental approval.
			6)	Planfulness	6) Tendency to plan ahead and to deliberate over decisions.
			7)	Rumination	7) Tendency to obsessively worry about past errors, less than perfect performance, or future mistakes.
			8)	Striving for Excellence	8) Tendency to pursue perfect results and high standards.
Almost Perfect Scale -Revised	Slaney et al., 2001	23	1)	Standards	1) Perfectionistic strivings and expectations of high performance.
			2)	Order	2) Preference for organization.
			3)	Discrepancy	3) Concern and perceived gap between personal standards and having met these.
Perfectionistic Self-Presentation Scale	Hewitt et al., 2003	27	1)	Perfectionistic self-promotion	1) Displaying one's perfection.
			2)	Non-display of imperfection	2) Concealing one's imperfection.
			3)	Non-disclosure of imperfection	3) Evading verbal admissions of one's imperfection.
Positive and Negative Perfectionism Scale	Terry-Short et al., 1995	40	1)	Positive	1) The motivation to achieve a goal and obtain a positive outcome.
			2)	Negative	2) The motivation to achieve a goal to avoid negative consequences.

Measure	Author(s)	No. of Items	Dimensions	Definition
Obsessive Beliefs Questionnaire – Perfectionism and need for Certainty subscale	Obsessive Compulsive Cognitions Working Group, 2005	16	Unidimensional	To assess dysfunctional beliefs thought to contribute to the development of OCD.
Neurotic Perfectionism Questionnaire	Mitzman, Slade, & Dewey, 1994	42	Unidimensional	The degree individuals set unrealistically high standards for themselves, are concerned about making mistakes, and are driven by a fear of failure.
Eating Disorders Inventory – Perfectionism subscale	Garner, 1991	6	Unidimensional	The belief that only the highest standards of personal performance are acceptable, and that outstanding achievement is expected by others.

## 6.2. Quality Assessment Tool for **Observational Cohort** and **Cross-Sectional Studies**

Criteria	Yes	No	Other (CD, NR, NA)*
1. Was the research question or objective in this paper clearly stated?			
2. Was the study population clearly specified and defined?			
3. Was the participation rate of eligible persons at least 50%?			
4. Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study pre- specified and applied uniformly to all participants?			
5. Was a sample size justification, power description, or variance and effect estimates provided?			
6. For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?			
7. Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?			
8. For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as continuous variable)?			
9. Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?			
10. Was the exposure(s) assessed more than once over time?			
11. Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?			
12. Were the outcome assessors blinded to the exposure status of participants?			
13. Was loss to follow-up after baseline 20% or less?			
14. Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)?			

\*CD = “cannot determine”; NA = “not applicable”; NR = “not reported”

Quality Rating (Good, Fair, or Poor) (see guidance)
Rater #1 initials:
Rater #2 initials:
Additional Comments (If POOR, please state why):

## **Guidance for Assessing the Quality of Observational Cohort and Cross-Sectional Studies**

The guidance document below is organized by question number from the tool for quality assessment of observational cohort and cross-sectional studies.

### **Question 1. Research question**

Did the authors describe their goal in conducting this research? Is it easy to understand what they were looking to find? This issue is important for any scientific paper of any type. Higher quality scientific research explicitly defines a research question.

### **Questions 2 and 3. Study population**

Did the authors describe the group of people from which the study participants were selected or recruited, using demographics, location, and time period? If you were to conduct this study again, would you know who to recruit, from where, and from what time period? Is the cohort population free of the outcomes of interest at the time they were recruited?

An example would be men over 40 years old with type 2 diabetes who began seeking medical care at Phoenix Good Samaritan Hospital between January 1, 1990 and December 31, 1994. In this example, the population is clearly described as: (1) who (men over 40 years old with type 2 diabetes); (2) where (Phoenix Good Samaritan Hospital); and (3) when (between January 1, 1990 and December 31, 1994). Another example is women ages 34 to 59 years of age in 1980 who were in the nursing profession and had no known coronary disease, stroke, cancer, hypercholesterolemia, or diabetes, and were recruited from the 11 most populous States, with contact information obtained from State nursing boards.

In cohort studies, it is crucial that the population at baseline is free of the outcome of interest. For example, the nurses' population above would be an appropriate group in which to study incident coronary disease. This information is usually found either in descriptions of population recruitment, definitions of variables, or inclusion/exclusion criteria.

You may need to look at prior papers on methods in order to make the assessment for this question. Those papers are usually in the reference list.

If fewer than 50% of eligible persons participated in the study, then there is concern that the study population does not adequately represent the target population. This increases the risk of bias.

#### **Question 4. Groups recruited from the same population and uniform eligibility criteria**

Were the inclusion and exclusion criteria developed prior to recruitment or selection of the study population? Were the same underlying criteria used for all of the subjects involved? This issue is related to the description of the study population, above, and you may find the information for both of these questions in the same section of the paper.

Most cohort studies begin with the selection of the cohort; participants in this cohort are then measured or evaluated to determine their exposure status. However, some cohort studies may recruit or select exposed participants in a different time or place than unexposed participants, especially retrospective cohort studies—which is when data are obtained from the past (retrospectively), but the analysis examines exposures prior to outcomes. For example, one research question could be whether diabetic men with clinical depression are at higher risk for cardiovascular disease than those without clinical depression. So, diabetic men with depression might be selected from a mental health clinic, while diabetic men without depression might be selected from an internal medicine or endocrinology clinic. This study recruits groups from different clinic populations, so this example would get a "no."

However, the women nurses described in the question above were selected based on the same inclusion/exclusion criteria, so that example would get a "yes."

#### **Question 5. Sample size justification**

Did the authors present their reasons for selecting or recruiting the number of people included or analysed? Do they note or discuss the statistical power of the study? This question is about whether or not the study had enough participants to detect an association if one truly existed.

A paragraph in the methods section of the article may explain the sample size needed to detect a hypothesized difference in outcomes. You may also find a discussion of power in the discussion section (such as the study had 85 percent power to detect a 20 percent increase in the rate of an outcome of interest, with a 2-sided alpha of 0.05). Sometimes estimates of variance and/or estimates of effect size are given, instead of sample size calculations. In any of these cases, the answer would be "yes."

However, observational cohort studies often do not report anything about power or sample sizes because the analyses are exploratory in nature. In this case, the answer would be "no." This is not a "fatal flaw." It just may indicate that attention was not paid to whether the study was sufficiently sized to answer a pre-specified question—i.e., it may have been an exploratory, hypothesis-generating study.



### **Question 6. Exposure assessed prior to outcome measurement**

This question is important because, in order to determine whether an exposure causes an outcome, the exposure must come before the outcome.

For some prospective cohort studies, the investigator enrolls the cohort and then determines the exposure status of various members of the cohort (large epidemiological studies like Framingham used this approach). However, for other cohort studies, the cohort is selected based on its exposure status, as in the example above of depressed diabetic men (the exposure being depression). Other examples include a cohort identified by its exposure to fluoridated drinking water and then compared to a cohort living in an area without fluoridated water, or a cohort of military personnel exposed to combat in the Gulf War compared to a cohort of military personnel not deployed in a combat zone.

With either of these types of cohort studies, the cohort is followed forward in time (i.e., prospectively) to assess the outcomes that occurred in the exposed members compared to non-exposed members of the cohort. Therefore, you begin the study in the present by looking at groups that were exposed (or not) to some biological or behavioural factor, intervention, etc., and then you follow them forward in time to examine outcomes. If a cohort study is conducted properly, the answer to this question should be "yes," since the exposure status of members of the cohort was determined at the beginning of the study before the outcomes occurred.

For retrospective cohort studies, the same principal applies. The difference is that, rather than identifying a cohort in the present and following them forward in time, the investigators go back in time (i.e., retrospectively) and select a cohort based on their exposure status in the past and then follow them forward to assess the outcomes that occurred in the exposed and non-exposed cohort members. Because in retrospective cohort studies the exposure and outcomes may have already occurred (it depends on how long they follow the cohort), it is important to make sure that the exposure preceded the outcome.

Sometimes cross-sectional studies are conducted (or cross-sectional analyses of cohort-study data), where the exposures and outcomes are measured during the same timeframe. As a result, cross-sectional analyses provide weaker evidence than regular cohort studies regarding a potential causal relationship between exposures and outcomes. For cross-sectional analyses, the answer to Question 6 should be "no."

### **Question 7. Sufficient timeframe to see an effect**

Did the study allow enough time for a sufficient number of outcomes to occur or be observed, or enough time for an exposure to have a biological effect on an outcome? In the examples given above, if clinical depression has a biological effect on increasing risk for CVD, such an effect may take years. In the other example, if higher dietary sodium increases BP, a short timeframe may be sufficient to assess its association with BP, but a longer timeframe would be needed to examine its association with heart attacks.

The issue of timeframe is important to enable meaningful analysis of the relationships between exposures and outcomes to be conducted. This often requires at least several years, especially when looking at health outcomes, but it depends on the research question and outcomes being examined.

Cross-sectional analyses allow no time to see an effect, since the exposures and outcomes are assessed at the same time, so those would get a "no" response.

#### **Question 8. Different levels of the exposure of interest**

If the exposure can be defined as a range (examples: drug dosage, amount of physical activity, amount of sodium consumed), were multiple categories of that exposure assessed? (for example, for drugs: not on the medication, on a low dose, medium dose, high dose; for dietary sodium, higher than average U.S. consumption, lower than recommended consumption, between the two). Sometimes discrete categories of exposure are not used, but instead exposures are measured as continuous variables (for example, mg/day of dietary sodium or BP values).

In any case, studying different levels of exposure (where possible) enables investigators to assess trends or dose-response relationships between exposures and outcomes—e.g., the higher the exposure, the greater the rate of the health outcome. The presence of trends or dose-response relationships lends credibility to the hypothesis of causality between exposure and outcome.

For some exposures, however, this question may not be applicable (e.g., the exposure may be a dichotomous variable like living in a rural setting versus an urban setting, or vaccinated/not vaccinated with a one-time vaccine). If there are only two possible exposures (yes/no), then this question should be given an "NA," and it should not count negatively towards the quality rating.

#### **Question 9. Exposure measures and assessment**

Were the exposure measures defined in detail? Were the tools or methods used to measure exposure accurate and reliable—for example, have they been validated or are they objective? This issue is important as it influences confidence in the reported exposures. When exposures are measured with less accuracy or validity, it is harder to see an association between exposure and outcome even if one exists. Also as important is whether the exposures were assessed in the same manner within groups and between groups; if not, bias may result.

For example, retrospective self-report of dietary salt intake is not as valid and reliable as prospectively using a standardized dietary log plus testing participants' urine for sodium content. Another example is measurement of BP, where there may be quite a difference between usual care, where clinicians measure BP however it is done in their practice setting (which can vary considerably), and use of trained BP assessors using standardized equipment (e.g., the same BP device which has been tested and calibrated) and a standardized protocol (e.g., patient is seated for 5 minutes with feet flat on the floor, BP is taken twice in each arm, and all four measurements are averaged). In each of these cases, the former would get a "no" and the latter a "yes."

Here is a final example that illustrates the point about why it is important to assess exposures consistently across all groups: If people with higher BP (exposed cohort) are seen by their providers more frequently than those without elevated BP (non-exposed group), it also increases the chances of detecting and documenting changes in health outcomes, including CVD-related events. Therefore, it may lead to the conclusion that higher BP leads to more CVD events. This may be true, but it could also be due to the fact that the subjects with higher BP were seen more often; thus, more CVD-related events were detected and documented simply because they had more encounters with the health care system. Thus, it could bias the results and lead to an erroneous conclusion.

### **Question 10. Repeated exposure assessment**

Was the exposure for each person measured more than once during the course of the study period? Multiple measurements with the same result increase our confidence that the exposure status was correctly classified. Also, multiple measurements enable investigators to look at changes in exposure over time, for example, people who ate high dietary sodium throughout the follow-up period, compared to those who started out high then reduced their intake, compared to those who ate low sodium throughout. Once again, this may not be applicable in all cases. In many older studies, exposure was measured only at baseline. However, multiple exposure measurements do result in a stronger study design.

### **Question 11. Outcome measures**

Were the outcomes defined in detail? Were the tools or methods for measuring outcomes accurate and reliable—for example, have they been validated or are they objective? This issue is important because it influences confidence in the validity of study results. Also important is whether the outcomes were assessed in the same manner within groups and between groups.

An example of an outcome measure that is objective, accurate, and reliable is death—the outcome measured with more accuracy than any other. But even with a measure as objective as death, there can be differences in the accuracy and reliability of how death was assessed by the investigators. Did they base it on an autopsy report, death certificate, death registry, or report from a family member? Another example is a study of whether dietary fat intake is related to blood cholesterol level (cholesterol level being the outcome), and the cholesterol level is measured from fasting blood samples that are all sent to the same laboratory. These examples would get a "yes." An example of a "no" would be self-report by subjects that they had a heart attack, or self-report of how much they weigh (if body weight is the outcome of interest).

Similar to the example in Question 9, results may be biased if one group (e.g., people with high BP) is seen more frequently than another group (people with normal BP) because more frequent encounters with the health care system increases the chances of outcomes being detected and documented.

### **Question 12. Blinding of outcome assessors**

Blinding means that outcome assessors did not know whether the participant was exposed or unexposed. It is also sometimes called "masking." The objective is to look for evidence in the article that the person(s) assessing the outcome(s) for the study (for example, examining medical records to determine the outcomes that occurred in the exposed and comparison groups) is masked to the exposure status of the participant. Sometimes the person measuring the exposure is the same person conducting the outcome assessment. In this case, the outcome assessor would most likely not be blinded to exposure status because they also took measurements of exposures. If so, make a note of that in the comments section.

As you assess this criterion, think about whether it is likely that the person(s) doing the outcome assessment would know (or be able to figure out) the exposure status of the study participants. If the answer is no, then blinding is adequate. An example of adequate blinding of the outcome assessors is to create a separate committee, whose members were not involved in the care of the patient and had no information about the study participants' exposure status. The committee

would then be provided with copies of participants' medical records, which had been stripped of any potential exposure information or personally identifiable information. The committee would then review the records for pre-specified outcomes according to the study protocol. If blinding was not possible, which is sometimes the case, mark "NA" and explain the potential for bias.

### **Question 13. Follow-up rate**

Higher overall follow-up rates are always better than lower follow-up rates, even though higher rates are expected in shorter studies, whereas lower overall follow-up rates are often seen in studies of longer duration. Usually, an acceptable overall follow-up rate is considered 80 percent or more of participants whose exposures were measured at baseline. However, this is just a general guideline. For example, a 6-month cohort study examining the relationship between dietary sodium intake and BP level may have over 90 percent follow-up, but a 20-year cohort study examining effects of sodium intake on stroke may have only a 65 percent follow-up rate.

### **Question 14. Statistical analyses**

Were key potential confounding variables measured and adjusted for, such as by statistical adjustment for baseline differences? Logistic regression or other regression methods are often used to account for the influence of variables not of interest.

This is a key issue in cohort studies, because statistical analyses need to control for potential confounders, in contrast to an RCT, where the randomization process controls for potential confounders. All key factors that may be associated both with the exposure of interest and the outcome—that are not of interest to the research question—should be controlled for in the analyses.

For example, in a study of the relationship between cardiorespiratory fitness and CVD events (heart attacks and strokes), the study should control for age, BP, blood cholesterol, and body weight, because all of these factors are associated both with low fitness and with CVD events. Well-done cohort studies control for multiple potential confounders.

### **Some general guidance for determining the overall quality rating of observational cohort and cross-sectional studies**

The questions on the form are designed to help you focus on the key concepts for evaluating the internal validity of a study. They are not intended to create a list that you simply tally up to arrive at a summary judgment of quality.

Internal validity for cohort studies is the extent to which the results reported in the study can truly be attributed to the exposure being evaluated and not to flaws in the design or conduct of the study—in other words, the ability of the study to draw associative conclusions about the effects of the exposures being studied on outcomes. Any such flaws can increase the risk of bias.

Critical appraisal involves considering the risk of potential for selection bias, information bias, measurement bias, or confounding (the mixture of exposures that one cannot tease out from each other). Examples of confounding include co-interventions, differences at baseline in patient characteristics, and other issues throughout the questions above. High risk of bias translates to a rating of poor quality. Low risk of bias translates to a rating of good quality. (Thus, the greater the risk of bias, the lower the quality rating of the study.)

In addition, the more attention in the study design to issues that can help determine whether there is a causal relationship between the exposure and outcome, the higher quality the study. These include exposures occurring prior to outcomes, evaluation of a dose-response gradient, and accuracy of measurement of both exposure and outcome, sufficient timeframe to see an effect, and appropriate control for confounding—all concepts reflected in the tool.

Generally, when you evaluate a study, you will not see a "fatal flaw," but you will find some risk of bias. By focusing on the concepts underlying the questions in the quality assessment tool, you should ask yourself about the potential for bias in the study you are critically appraising. For any box where you check "no" you should ask, "What is the potential risk of bias resulting from this flaw in study design or execution?" That is, does this factor cause you to doubt the results that are reported in the study or doubt the ability of the study to accurately assess an association between exposure and outcome?

The best approach is to think about the questions in the tool and how each one tells you something about the potential for bias in a study. The more you familiarize yourself with the key concepts, the more comfortable you will be with critical appraisal. Examples of studies rated good, fair, and poor are useful, but each study must be assessed on its own based on the details that are reported and consideration of the concepts for minimizing bias.

### 6.3. Quality Assessment Tool for Case-Control Studies.

Criteria	Yes	No	Other (CD, NR, NA)*
1. Was the research question or objective in this paper clearly stated and appropriate?			
2. Was the study population clearly specified and defined?			
3. Did the authors include a sample size justification?			
4. Were controls selected or recruited from the same or similar population that gave rise to the cases (including the same timeframe)?			
5. Were the definitions, inclusion and exclusion criteria, algorithms or processes used to identify or select cases and controls valid, reliable, and implemented consistently across all study participants?			
6. Were the cases clearly defined and differentiated from controls?			
7. If less than 100 percent of eligible cases and/or controls were selected for the study, were the cases and/or controls randomly selected from those eligible?			
8. Was there use of concurrent controls?			
9. Were the investigators able to confirm that the exposure/risk occurred prior to the development of the condition or event that defined a participant as a case?			
10. Were the measures of exposure/risk clearly defined, valid, reliable, and implemented consistently (including the same time period) across all study participants?			
11. Were the assessors of exposure/risk blinded to the case or control status of participants?			
12. Were key potential confounding variables measured and adjusted statistically in the analyses? If matching was used, did the investigators account for matching during study analysis?			

\*CD = "cannot determine"; NA = "not applicable"; NR = "not reported"

Quality Rating (Good, Fair, or Poor) (see guidance)
Rater #1 initials:
Rater #2 initials:
Additional Comments (If POOR, please state why):

## **Guidance for Assessing the Quality of Case-Control Studies**

*The guidance document below is organized by question number from the tool for quality assessment of case-control studies.*

### **Question 1. Research question**

Did the authors describe their goal in conducting this research? Is it easy to understand what they were looking to find? This issue is important for any scientific paper of any type. High quality scientific research explicitly defines a research question.

### **Question 2. Study population**

Did the authors describe the group of individuals from which the cases and controls were selected or recruited, while using demographics, location, and time period? If the investigators conducted this study again, would they know exactly who to recruit, from where, and from what time period?

Investigators identify case-control study populations by location, time period, and inclusion criteria for cases (individuals with the disease, condition, or problem) and controls (individuals without the disease, condition, or problem). For example, the population for a study of lung cancer and chemical exposure would be all incident cases of lung cancer diagnosed in patients ages 35 to 79, from January 1, 2003 to December 31, 2008, living in Texas during that entire time period, as well as controls without lung cancer recruited from the same population during the same time period. The population is clearly described as: (1) who (men and women ages 35 to 79 with (cases) and without (controls) incident lung cancer); (2) where (living in Texas); and (3) when (between January 1, 2003 and December 31, 2008).

Other studies may use disease registries or data from cohort studies to identify cases. In these cases, the populations are individuals who live in the area covered by the disease registry or included in a cohort study (i.e., nested case-control or case-cohort). For example, a study of the relationship between vitamin D intake and myocardial infarction might use patients identified via the GRACE registry, a database of heart attack patients.

NHLBI staff encouraged reviewers to examine prior papers on methods (listed in the reference list) to make this assessment, if necessary.

### **Question 3i. Target population and case representation**

In order for a study to truly address the research question, the target population—the population from which the study population is drawn and to which study results are believed to apply—should be carefully defined. Some authors may compare characteristics of the study cases to characteristics of cases in the target population, either in text or in a table. When study cases are shown to be representative of cases in the appropriate target population, it increases the likelihood that the study was well-designed per the research question.

However, because these statistics are frequently difficult or impossible to measure, publications should not be penalized if case representation is not shown. For most papers, the response to

question 3 will be "NR." Those subquestions are combined because the answer to the second subquestion—case representation—determines the response to this item. However, it cannot be determined without considering the response to the first subquestion. For example, if the answer to the first subquestion is "yes," and the second, "CD," then the response for item 3 is "CD."

### **Question 3ii. Sample size justification**

Did the authors discuss their reasons for selecting or recruiting the number of individuals included? Did they discuss the statistical power of the study and provide a sample size calculation to ensure that the study is adequately powered to detect an association (if one exists)? This question does not refer to a description of the manner in which different groups were included or excluded using the inclusion/exclusion criteria (e.g., "Final study size was 1,378 participants after exclusion of 461 patients with missing data" is not considered a sample size justification for the purposes of this question).

An article's methods section usually contains information on sample size and the size needed to detect differences in exposures and on statistical power.

### **Question 4. Groups recruited from the same population**

To determine whether cases and controls were recruited from the same population, one can ask hypothetically, "If a control was to develop the outcome of interest (the condition that was used to select cases), would that person have been eligible to become a case?" Case-control studies begin with the selection of the cases (those with the outcome of interest, e.g., lung cancer) and controls (those in whom the outcome is absent). Cases and controls are then evaluated and categorized by their exposure status. For the lung cancer example, cases and controls were recruited from hospitals in a given region. One may reasonably assume that controls in the catchment area for the hospitals, or those already in the hospitals for a different reason, would attend those hospitals if they became a case; therefore, the controls are drawn from the same population as the cases. If the controls were recruited or selected from a different region (e.g., a State other than Texas) or time period (e.g., 1991-2000), then the cases and controls were recruited from different populations, and the answer to this question would be "no."

The following example further explores selection of controls. In a study, eligible cases were men and women, ages 18 to 39, who were diagnosed with atherosclerosis at hospitals in Perth, Australia, between July 1, 2000 and December 31, 2007. Appropriate controls for these cases might be sampled using voter registration information for men and women ages 18 to 39, living in Perth (population-based controls); they also could be sampled from patients without atherosclerosis at the same hospitals (hospital-based controls). As long as the controls are individuals who would have been eligible to be included in the study as cases (if they had been diagnosed with atherosclerosis), then the controls were selected appropriately from the same source population as cases.

In a prospective case-control study, investigators may enrol individuals as cases at the time they are found to have the outcome of interest; the number of cases usually increases as time progresses. At this same time, they may recruit or select controls from the population without the outcome of interest. One way to identify or recruit cases is through a surveillance system. In turn, investigators can select controls from the population covered by that system. This is an example of population-based controls. Investigators also may identify and select cases from a cohort study



population and identify controls from outcome-free individuals in the same cohort study. This is known as a nested case-control study.

#### **Question 5. Inclusion and exclusion criteria prespecified and applied uniformly**

Were the inclusion and exclusion criteria developed prior to recruitment or selection of the study population? Were the same underlying criteria used for all of the groups involved? To answer this question, reviewers determined if the investigators developed I/E criteria prior to recruitment or selection of the study population and if they used the same underlying criteria for all groups. The investigators should have used the same selection criteria, except for study participants who had the disease or condition, which would be different for cases and controls by definition. Therefore, the investigators use the same age (or age range), gender, race, and other characteristics to select cases and controls. Information on this topic is usually found in a paper's section on the description of the study population.

#### **Question 6. Case and control definitions**

For this question, reviewers looked for descriptions of the validity of case and control definitions and processes or tools used to identify study participants as such. Was a specific description of "case" and "control" provided? Is there a discussion of the validity of the case and control definitions and the processes or tools used to identify study participants as such? They determined if the tools or methods were accurate, reliable, and objective. For example, cases might be identified as "adult patients admitted to a VA hospital from January 1, 2000 to December 31, 2009, with an ICD-9 discharge diagnosis code of acute myocardial infarction and at least one of the two confirmatory findings in their medical records: at least 2mm of ST elevation changes in two or more ECG leads and an elevated troponin level. Investigators might also use ICD-9 or CPT codes to identify patients. All cases should be identified using the same methods. Unless the distinction between cases and controls is accurate and reliable, investigators cannot use study results to draw valid conclusions.

#### **Question 7. Random selection of study participants**

If a case-control study did not use 100 percent of eligible cases and/or controls (e.g., not all disease-free participants were included as controls), did the authors indicate that random sampling was used to select controls? When it is possible to identify the source population fairly explicitly (e.g., in a nested case-control study, or in a registry-based study), then random sampling of controls is preferred. When investigators used consecutive sampling, which is frequently done for cases in prospective studies, then study participants are not considered randomly selected. In this case, the reviewers would answer "no" to Question 8. However, this would not be considered a fatal flaw.

If investigators included all eligible cases and controls as study participants, then reviewers marked "NA" in the tool. If 100 percent of cases were included (e.g., NA for cases) but only 50 percent of eligible controls, then the response would be "yes" if the controls were randomly selected, and "no" if they were not. If this cannot be determined, the appropriate response is "CD."

### **Question 8. Concurrent controls**

A concurrent control is a control selected at the time another person became a case, usually on the same day. This means that one or more controls are recruited or selected from the population without the outcome of interest at the time a case is diagnosed. Investigators can use this method in both prospective case-control studies and retrospective case-control studies. For example, in a retrospective study of adenocarcinoma of the colon using data from hospital records, if hospital records indicate that Person A was diagnosed with adenocarcinoma of the colon on June 22, 2002, then investigators would select one or more controls from the population of patients without adenocarcinoma of the colon on that same day. This assumes they conducted the study retrospectively, using data from hospital records. The investigators could have also conducted this study using patient records from a cohort study, in which case it would be a nested case-control study.

Investigators can use concurrent controls in the presence or absence of matching and vice versa. A study that uses matching does not necessarily mean that concurrent controls were used.

### **Question 9. Exposure assessed prior to outcome measurement**

Investigators first determine case or control status (based on presence or absence of outcome of interest), and then assess exposure history of the case or control; therefore, reviewers ascertained that the exposure preceded the outcome. For example, if the investigators used tissue samples to determine exposure, did they collect them from patients prior to their diagnosis? If hospital records were used, did investigators verify that the date a patient was exposed (e.g., received medication for atherosclerosis) occurred prior to the date they became a case (e.g., was diagnosed with type 2 diabetes)? For an association between an exposure and an outcome to be considered causal, the exposure must have occurred prior to the outcome.

### **Question 10. Exposure measures and assessment**

Were the exposure measures defined in detail? Were the tools or methods used to measure exposure accurate and reliable—for example, have they been validated or are they objective? This is important, as it influences confidence in the reported exposures. Equally important is whether the exposures were assessed in the same manner within groups and between groups. This question pertains to bias resulting from exposure misclassification (i.e., exposure ascertainment).

For example, a retrospective self-report of dietary salt intake is not as valid and reliable as prospectively using a standardized dietary log plus testing participants' urine for sodium content because participants' retrospective recall of dietary salt intake may be inaccurate and result in misclassification of exposure status. Similarly, BP results from practices that use an established protocol for measuring BP would be considered more valid and reliable than results from practices that did not use standard protocols. A protocol may include using trained BP assessors, standardized equipment (e.g., the same BP device which has been tested and calibrated), and a standardized procedure (e.g., patient is seated for 5 minutes with feet flat on the floor, BP is taken twice in each arm, and all four measurements are averaged).

### **Question 11. Blinding of exposure assessors**

Blinding or masking means that outcome assessors did not know whether participants were exposed or unexposed. To answer this question, reviewers examined articles for evidence that the outcome assessor(s) was masked to the exposure status of the research participants. An outcome assessor, for example, may examine medical records to determine the outcomes that occurred in the exposed and comparison groups. Sometimes the person measuring the exposure is the same person conducting the outcome assessment. In this case, the outcome assessor would most likely not be blinded to exposure status. A reviewer would note such a finding in the comments section of the assessment tool.

One way to ensure good blinding of exposure assessment is to have a separate committee, whose members have no information about the study participants' status as cases or controls, review research participants' records. To help answer the question above, reviewers determined if it was likely that the outcome assessor knew whether the study participant was a case or control. If it was unlikely, then the reviewers marked "no" to Question 12. Outcome assessors who used medical records to assess exposure should not have been directly involved in the study participants' care, since they probably would have known about their patients' conditions. If the medical records contained information on the patient's condition that identified him/her as a case (which is likely), that information would have had to be removed before the exposure assessors reviewed the records.

If blinding was not possible, which sometimes happens, the reviewers marked "NA" in the assessment tool and explained the potential for bias.

### **Question 12. Statistical analysis**

Were key potential confounding variables measured and adjusted for, such as by statistical adjustment for baseline differences? Investigators often use logistic regression or other regression methods to account for the influence of variables not of interest.

This is a key issue in case-controlled studies; statistical analyses need to control for potential confounders, in contrast to RCTs in which the randomization process controls for potential confounders. In the analysis, investigators need to control for all key factors that may be associated with both the exposure of interest and the outcome and are not of interest to the research question.

A study of the relationship between smoking and CVD events illustrates this point. Such a study needs to control for age, gender, and body weight; all are associated with smoking and CVD events. Well-done case-control studies control for multiple potential confounders.

Matching is a technique used to improve study efficiency and control for known confounders. For example, in the study of smoking and CVD events, an investigator might identify cases that have had a heart attack or stroke and then select controls of similar age, gender, and body weight to the cases. For case-control studies, it is important that if matching was performed during the selection or recruitment process, the variables used as matching criteria (e.g., age, gender, race) should be controlled for in the analysis.

## **General Guidance for Determining the Overall Quality Rating of Case-Controlled Studies**

NHLBI designed the questions in the assessment tool to help reviewers focus on the key concepts for evaluating a study's internal validity, not to use as a list from which to add up items to judge a study's quality.

Internal validity for case-control studies is the extent to which the associations between disease and exposure reported in the study can truly be attributed to the exposure being evaluated rather than to flaws in the design or conduct of the study. In other words, what is ability of the study to draw associative conclusions about the effects of the exposures on outcomes? Any such flaws can increase the risk of bias.

In critical appraising a study, the following factors need to be considered: risk of potential for selection bias, information bias, measurement bias, or confounding (the mixture of exposures that one cannot tease out from each other). Examples of confounding include co-interventions, differences at baseline in patient characteristics, and other issues addressed in the questions above. High risk of bias translates to a poor quality rating; low risk of bias translates to a good quality rating. Again, the greater the risk of bias, the lower the quality rating of the study.

In addition, the more attention in the study design to issues that can help determine whether there is a causal relationship between the outcome and the exposure, the higher the quality of the study. These include exposures occurring prior to outcomes, evaluation of a dose-response gradient, accuracy of measurement of both exposure and outcome, sufficient timeframe to see an effect, and appropriate control for confounding—all concepts reflected in the tool.

If a study has a "fatal flaw," then risk of bias is significant; therefore, the study is deemed to be of poor quality. An example of a fatal flaw in case-control studies is a lack of a consistent standard process used to identify cases and controls.

Generally, when reviewers evaluated a study, they did not see a "fatal flaw," but instead found some risk of bias. By focusing on the concepts underlying the questions in the quality assessment tool, reviewers examined the potential for bias in the study. For any box checked "no," reviewers asked, "What is the potential risk of bias resulting from this flaw in study design or execution?" That is, did this factor lead to doubt about the results reported in the study or the ability of the study to accurately assess an association between exposure and outcome?

By examining questions in the assessment tool, reviewers were best able to assess the potential for bias in a study. Specific rules were not useful, as each study had specific nuances. In addition, being familiar with the key concepts helped reviewers assess the studies. Examples of studies rated good, fair, and poor were useful, yet each study had to be assessed on its own.

#### 6.4. Table summarising the appearance-related outcome measures used by the studies.

Measure	Author(s)	No. of Items	Construct	Direction of maladaptive scores
Aging-related Concerns about Appearance	Gupta & Schork, 1993	5	Concerns about aging appearance and body image.	↑
Appearance Anxiety Inventory	Veale et al., 2014	10	Symptoms of body dysmorphic disorder.	↑
Appearance Schemas Inventory	Cash & Labarge, 1996	14	Core beliefs/assumptions about the importance, meaning, and effects of physical appearance in one's life.	↑
Assessment of Body Image Cognitive Distortions	Jakatdar, Cash & Engle, 2006	37	Frequency of cognitive errors or distortions related to body-image thoughts in different situations.	↑
BDD Modification of the Yale-Brown OCD Scale	Phillips et al., 1997	12	Severity of body dysmorphic disorder symptoms during the past week.	↑
Body Change Inventory – Muscle Preoccupation subscale	Ricciardelli & McCabe, 2002	8	Frequency of strategies used by both adolescent girls and boys to increase muscle size.	↑
Body Dysmorphic Disorder Questionnaire	Phillips, 2005	4	Screening instrument for Body Dysmorphic Disorder based on the criteria outlined in DSM-IV.	NA - categorical
Body Esteem Scale	Franzoi & Shields, 1984	35	Attitudes towards body parts and functions.	↓
Body Esteem Scale for Adolescents and Adults	Mendelson, Mendelson, & White, 2001	23	Attitudes towards one's appearance, weight, and attribution (evaluations attributed to others about one's body and appearance).	↓
Body Image Concerns Inventory	Littleton, Axsom & Pury, 2005	19	Dysmorphic concerns about appearance.	↑
Body Image Disturbance Questionnaire	Cash, Phillips, Santos & Hrabosky, 2004	7	Body image dissatisfaction, distress (or dysphoria), and dysfunction (or impairment).	↑

Measure	Author(s)	No. of Items	Construct	Direction of maladaptive scores
Body Image Ideals Questionnaire	Cash & Szymanski, 1995	20	Evaluation of 10 physical characteristics which includes the individual's perceived discrepancy from ideal (discrepancy score), the importance attributed to this (importance score), and a weighted score incorporating both (weighted-discrepancy).	↑
Body Image Rating Scale	Mayville, Katz, Gipson & Cabral, 1999	14	Frequency and intensity of distress, impairment, and dissatisfaction tied to appearance-related concerns.	↑
Body Image Satisfaction Scale	Holsen, Jones, & Birkeland, 2012	4	Satisfaction with body and appearance.	↓
Body Image States Scale	Cash, 2007; Cash et al., 2002	6	Body-image experiences at a particular point in time or in a specific context.	↓
Body Image Test	Jade, 2002 (cited in Malkina-Pykh, 2012)	20	Concerns with appearance in different situations.	↑
Body Parts Satisfaction Scale	Berscheid, Walster, & Bohrnstedt, 1973	24	Satisfaction with body image by focusing on specific body parts and features.	↓
Body Parts Satisfaction Scale – Revised	Petrie, Tripp & Harvey, 2002	11	Satisfaction with body image by focusing on specific body parts (7 items) and facial features (4 items).	↓
Body Self-Relations Questionnaire	Thompson, Heinberg, Altabe & Tantleff-Dunn, 1999 (cited in Cooley et al., 2007)	7	Measure of the participant's personal satisfaction with their physical appearance.	CD
Contingencies of Self Worth Scale – Appearance Subscale	Crocker, Luhtanen, Cooper & Bouvrette, 2003	5	Degree to which self-worth was perceived as contingent on appearance.	↑

Measure	Author(s)	No. of Items	Construct	Direction of maladaptive scores
Drive for Muscularity Scale	McCreary & Sasse, 2000	15	There are two subscales: one reflecting body-image attitudes concerning one's muscularity (Attitudes – 8-items), and the other body-image behaviours that reflect one's desire to become larger and more muscular (Behaviours – 7-items).	↑
Dysmorphic Concerns Questionnaire	Oosthuizen, Lambert & Castle, 1998	7	Dysmorphic concerns about appearance.	↑
Male Body Attitudes Scale	Tylka, Bergeron & Schwartz, 2005	29	Men's attitudes toward their body and muscularity.	↑
Male Body Checking Questionnaire	Hildebrandt, Walker, Alfano, Delinsky & Bannon, 2010	19	The extent men regularly check their muscularity, an important behavioural consequence of body image disturbance.	↑
Male Body Dissatisfaction Scale	Hallsworth, Wade & Tiggemann, 2005 (validated in Italian by Dakanalis et al., 2015 <sub>b</sub> )	9	Male body dissatisfaction.	↑
Multidimensional Body-Self Relations Questionnaire – Appearance Scales	Brown, Cash, Mikulka, 1990	34	Attitudes towards body-image. It has five subscales, of which three were utilised by studies in this review. That is, appearance evaluation (AE; 7-items; satisfaction with one's looks); Appearance Orientation (AO; 12-items; extent of investment in one's appearance i.e. paying attention and engaging in extensive grooming behaviours); Body Areas Satisfaction Scale (BASS; 9-items; satisfaction with discrete aspects of one's appearance).	↓ (AE)
				↑ (AO)
				↓ (BASS)
Muscle Dysmorphia Disorder Inventory	Hildebrandt, Langenbucher, & Schlundt, 2004 (Validated in Italian by Santaronechi & Dèttore, 2012)	13	Symptoms of muscle dysmorphia across three core features: Drive for Size, Appearance Intolerance, and Functional Impairment.	↑

Measure	Author(s)	No. of Items	Construct	Direction of maladaptive scores
Muscle Dysmorphia Questionnaire	Dryer, Farr, Hiramatsu & Quinton, 2016	34	Symptoms of muscle dysmorphia.	↑
Physical Self-Description Questionnaire – Appearance Subscale	Marsh, Richards, Johnson, Roche, & Tremayne, 1994	8	Perceptions of physical attractiveness, how appearance compares with others, and how others think they look.	↓
Self-Esteem Scale (Appearance subscale)	Heatherton & Polivy, 1991	6	Evaluation of appearance and body image.	↓
Social Appearance Anxiety Scale	Hart et al., 2008	16	Anxiety about being negatively evaluated by others because of one's overall appearance.	↑
Yelland & Tiggemann's Drive for Muscularity Scale	Yelland & Tiggemann, 2003	7	Motivation and engagement in behaviours designed to increase muscularity.	↑

*Note:* ↑ = higher scores on this measure are maladaptive; ↓ = lower scores on this measure are maladaptive; CD = Cannot Determine; and NA = Not Applicable



## MAIN EMPIRICAL PROJECT

---

# **Understanding “normal” appearance concerns in young people and how they are related to perfectionism:**

A cross-sectional study in a British school sample

---

**First Supervisor:** Dr Amita Jassi

**Second Supervisor:** Dr Georgina Krebs

## ABSTRACT

---

Appearance-related concerns are known to be common in the population and can develop into clinically significant worries as seen in Body Dysmorphic Disorder (BDD). Perfectionism, a multidimensional construct, is also theorised to play a role in the development and maintenance of such concerns. Yet neither normative concerns nor the role of perfectionism have been explored using a sample of British young people, despite evidence demonstrating that appearance-related concerns vary across ethnicity. The aims of the current study were two-fold; (1) to collect normative data on the extent to which adolescents are concerned about their appearance, and (2) explore the relationship between perfectionism and appearance-related concerns. A cross-sectional school study design was adopted to meet both aims. Firstly, students aged between 11 and 16 years ( $N = 534$ ) completed a questionnaire measuring BDD symptoms. Findings showed that 80% of young British people were concerned with at least one body part and on average checked their appearance up to five times a day, although concerns were associated with little distress, avoidance, or interference. Girls were also significantly more likely to report appearance-related concerns compared to boys, but concerns did not differ significantly by age. Secondly, a sub-sample of students aged 14-16 ( $N = 302$ ) completed questionnaires measuring BDD symptoms, perfectionism, and symptoms of anxiety/depression. Perfectionism was found to be significantly associated with appearance-related concerns. The self-oriented perfectionism dimension (but not socially-prescribed perfectionism) was found to be a unique predictor of the variance in appearance-related concerns whilst controlling for gender, anxiety, and depression. Therefore, perfectionism appears to play an important role in normative appearance-related concerns and should continue to be included in psychological models and treatment of BDD. This data can further be used in the assessment of BDD, but also in the education of non-clinical groups and professionals within the UK. Further research is warranted to replicate and advance our understanding in this area.

## CONTENTS

---

<b>1. INTRODUCTION .....</b>	<b>92</b>
1.1. Normative Appearance-Related Concerns.....	92
1.2. Perfectionism and Body Dysmorphic Disorder .....	95
1.3. The Current Study .....	99
<b>2. METHODS .....</b>	<b>100</b>
2.1. Design.....	100
2.2. Sample.....	100
2.3. Measures.....	102
2.3.1. Body Image Questionnaire – Child and Adolescent version (BIQ-C).....	102
2.4. Design.....	102
2.5. Sample.....	102
2.6. Measures.....	104
2.6.1. BIQ-C.....	104
2.6.2. Child-Adolescent Perfectionism Scale (CAPS) .....	104
2.6.3. Revised Child Anxiety and Depression Scale – Short Version (RCADS-25) .....	105
2.7. Procedure .....	106
2.8. Ethics .....	107
2.9. Data Analytic Plan .....	107
2.10. Power Analysis .....	108
<b>3. RESULTS.....</b>	<b>109</b>
3.1. Data Screening .....	109
3.2. Descriptive Statistics .....	109
3.2.1. Prevalence and area of appearance-related concerns.....	109
3.2.2. Time spent on Appearance-related Concerns.....	111
3.2.3. Quality of Appearance-related Concerns .....	112
3.3. Gender and Age Differences in Appearance-related Concerns .....	114
3.4. Descriptive Statistics .....	115
3.4.1. Correlations .....	116
3.5. Regression Analysis .....	117
3.5.1. Total Perfectionism Score.....	117
3.5.2. SOP/SPP Dimensions of Perfectionism.....	118
<b>4. DISCUSSION .....</b>	<b>120</b>
4.1. Summary – Study One.....	120
4.2. Summary – Study Two.....	122
4.3. General Limitations .....	124

4.4. Implications & Future Directions .....	125
4.5. Conclusion .....	127
<b>5. REFERENCES .....</b>	<b>129</b>
<b>6. APPENDIX .....</b>	<b>138</b>
6.1. Ethnic Composition of Sample from Study One.....	138
6.2. Body Image Questionnaire – Child and Adolescent version (BIQ-C; Veale, 2009).....	139
6.3. Ethnic Composition of Sample from Study Two.....	144
6.4. Child-Adolescent Perfectionism Scale (CAPS) .....	145
6.5. Revised Child Anxiety and Depression Scale – Short Version (RCADS-25).....	147
6.6. Ethical Approval Letter .....	149
6.7. Supplementary Statistical Analysis.....	150

# 1. INTRODUCTION

---

It is widely accepted that most people believe they have some imperfections in their appearance, however it is the perception of and responses to such appearance-related concerns that separates out those individuals with Body Dysmorphic Disorder (BDD) (Wilhelm & Neziroglu, 2002). BDD is characterised by a preoccupation with a perceived defect/ flaw in appearance that is not observable to others, and typically develops during adolescence (American Psychiatric Association, 2013; Phillips, 2001). BDD is distinct from “normal” appearance-related concerns as it is accompanied with substantial emotional distress and significant impairment that interferes with the individual’s everyday functioning. For instance, such preoccupations can consume many hours of the day, whereby the person may engage in ritualistic behaviours to compensate for the perceived defect (e.g. camouflaging, hiding, or checking), or will completely avoid potential evaluation by others (Fang, Matheny & Wilhelm, 2014). Compared to adults, adolescents with BDD have demonstrated significantly greater levels of distress and functional impairment (i.e. more delusional beliefs and suicide attempts), yet this usually goes unreported to their mental health providers (Phillips, Didie, Menard, Pagano, Fay & Weisberg, 2006). In adolescents, the prevalence of BDD has been estimated to range from 1.7-2.2% and does not appear to differ across gender, however, it is more prevalent in older compared to younger adolescents (Mayville, Katz, Gipson, & Cabral, 1999; Schneider, Turner, Mond, & Hudson, 2016<sub>a</sub>; Veale, Gledhill, Christodouloua, & Hodsoll, 2016). More recently Möllmann, Dietel, Hunger, and Buhlmann (2017) reviewed these figures according to the new DSM-V criteria and calculated a point prevalence rate of 3.6%.

## 1.1. Normative Appearance-Related Concerns

Appearance-related concerns can be conceptualised as being on a continuum, where BDD would be situated on the more severe end (Lambrou, Veale, & Wilson, 2012; Phillips, 2005). Many individuals typically fall within the middle of this spectrum, with appearance-related distress being in the mild to moderate range, but considered non-pathological (Tylka, 2012). However, it is unclear to what extent healthy and “normal” appearance-related concerns are present in young people, and when this dissatisfaction becomes a disturbance that is pervasive and interferes with day-to-day functioning (Cash, Phillips, Santos, & Hrabosky, 2004).

In adults, appearance-related concerns have been shown to be common. For example, 88% of adult controls reported at least one body feature that they disliked or felt was defective

(Lambrou et al., 2012). Harris and Carr (2001) further demonstrated that 35% of British men and 61% of British women expressed appearance-related concerns. Women were most frequently dissatisfied with their breasts, hair, skin, stomach, and nose, and men were mainly concerned with their hair, ears, and nose (Rief, Buhlmann, Wilhelm, Borkenhagen, & Brähler, 2006). Regarding samples of young people, few studies exist, although almost half (49.5%) of Australian school students (aged 12-18 years) were found to report some appearance-related concern (Schneider, Baillie, Mond, Turner, & Hudson, 2016<sub>b</sub>). Typically, the most common areas of concern amongst adolescents appear to be the skin, chest/breasts, nose, legs, stomach, hair, and weight/muscularity (Möllmann et al., 2017; Phillips et al., 2006; Schneider et al., 2017<sub>b</sub>).

Although a gender difference does not exist within clinical samples of BDD, when measures of BDD symptomology are given to school samples of adolescents in separate cross-sectional studies, girls are repeatedly found to report significantly more symptoms compared to boys (Mayville et al., 1999; Webb et al., 2015; Zimmer-Gembeck, Webb, Farrell & Waters, 2017). One group of researchers elaborated on this further by investigating gender differences in community adolescents who expressed “subthreshold-BDD” concerns compared to those indicated “probable-BDD” (Schneider et al., 2016<sub>a</sub>; Schneider, Mond, Turner, Hudson, 2017<sub>a</sub>). Sub-threshold BDD was more prevalent in girls compared to boys, whereas there were no gender differences identified within the probable-BDD group. Taken together, these studies suggest that normative appearance-related concerns are more common amongst female adolescents, although when the concerns become clinically significant the gender difference dissipates.

More specifically, this same research group further evaluated the presence of gender differences within the individual symptoms associated with BDD using those adolescents they had identified in the subthreshold and probable BDD groups (Schneider, Mond, Turner & Hudson, 2017<sub>b</sub>). Boys were found to report that their concerns interfered significantly more with socialising/dating and school work compared to girls. Girls were also found to report significantly more body areas of concern, although these differences did not remain significant after controlling for age, socio-educational advantage, and parent cultural background, or when correcting for multiple comparisons. Significant gender differences were observed for seven of 20 body areas analysed. Those concerning muscularity (greater in boys), breasts/nipples, and thighs (both greater in girls) remained significant after controlling for age, socio-educational advantage, and parent cultural background, whereas concerns about the chest, stomach, eyes, and teeth did not.

Age has additionally been identified as a factor associated with appearance-related concern, where older adolescents are more likely to express concerns compared to younger adolescents (Schneider et al., 2016<sub>a</sub>) – which is supported by the evidence suggesting that the onset of BDD is around 16 years of age (Phillips, 2001). Moreover, Rosenblum and Lewis (1999) found an interaction between gender and age on dissatisfaction towards body features. For instance, no significant gender differences were noted at age 13, however at age 15 girls' body dissatisfaction had increased, whereas the boys' dissatisfaction decreased. This difference was then maintained from age 15 to 18 years. Therefore, gender appears to only influence appearance-related concerns in older compared to younger adolescents. This has been further supported with longitudinal evidence that followed-up 387 Australian adolescents over two and a half years (Zimmer-Gembeck et al., 2017). Appearance anxiety was found to increase linearly over time in both genders, although for girls (and not boys), older adolescents expressed significantly more anxiety about their appearance relative to younger adolescents.

In addition to age and gender, appearance-related concerns have been shown to vary across ethnicity in adult university samples. For example, White American women reported significantly greater body image disturbance compared to African American women (Cash et al., 2004); non-clinical African American women reported significantly less symptoms of BDD compared to Latinas or Caucasian Americans, although no significant differences were identified for men (Boroughs, Krawczyk & Thompson, 2010); and Asian participants reported significantly less dysmorphic concerns compared to an Australian and European reference group (Bartsch, 2007). Moreover, University students from the USA ( $N = 101$ , 82.2% female,  $M$  age = 21 years) have been identified as reporting significantly more body image concerns and preoccupation compared to those from Germany ( $N = 133$ , 73.7% female,  $M$  age = 22 years), although these differences dissipated when comparing individuals who were likely to meet criteria for BDD (Bohne, Keuthen, Wilhelm, Deckersbach, & Jenike, 2002).

Phillips (2004) would argue that this was expected because clinical BDD symptoms have been observed to be similar across the world, however it is the expression of the BDD that varies and likely to be influenced by cultural norms. For instance, a Japanese case series reported BDD preoccupations of the eyelids, which is not typical of Western cultures, yet the features of the BDD (e.g. distress and impairment) were described to be comparable (Phillips, 2004). This phenomenon was evidenced in a community study, where Asian participants reported to be twice as likely to indicate concerns with their hair, and four times more likely to indicate concerns with their skin compared to Caucasians, and reported an increase in grooming, touching, camouflaging, and excessive exercise behaviours (Marques et al., 2011). Despite the

increasingly emerging evidence for cultural differences in BDD/appearance-related concern for adults, this has been a neglected area in adolescents. To our knowledge, one school study exists which identified a main effect of ethnicity on features of BDD in 566 American school children (51.1% male, *M* age = 15.8 year) (Mayville et al., 1999). Specifically, African Americans of both genders reported significantly less dissatisfaction with their bodies compared to Caucasians.

There have been no studies conducted to specifically explore appearance-related concerns in British adolescents, despite evidence suggesting appearance-related concerns vary across ethnicity, and so this is warranted. Because of this lack of research using British adolescent samples, and the large variation in samples and methods used in the available literature, it is difficult to predict what may be observed in the current study. Based on the literature outlined above, it would be hypothesised that with regards to self-reported appearance-related concerns, differences will be observed between British young people of different genders and ages. That is, girls and older children will be more likely to report appearance-related concerns.

## 1.2. Perfectionism and Body Dysmorphic Disorder

Perfectionism is not a construct exclusive to BDD. From clinical observations, Veale (2004) suggested that perfectionism may be one of the factors indirectly associated with the development and maintenance of BDD due to this temperament being somewhat genetically determined. This could have an evolutionary origin as perfectionist standards can be considered adaptive. This is because flaws and asymmetries are arguably seen to be indicative of poor genetic and physical health, thus interfering with mating success (Wilhelm and Naziroglu, 2002). Generally, cognitive-behavioural models of BDD suggest that such individuals interpret ordinary visual appearance-related stimuli in a maladaptive manner, which is influenced by cognitive biases and distortions. This encourages the individual to: selectively attend to their “flaws” or imperfections; process themselves as an aesthetic object; and/or engage in further negative mental, emotional, and behavioural consequences (Fang & Wilhelm, 2015; Veale, 2004; Veale & Neziroglu, 2010; Wilhelm & Neziroglu, 2002; Wilhelm, Phillips, & Steketee, 2012). Setting rigid perfectionistic standards for oneself can then drive the development of and/or combine with these interpretation biases and cognitive processes to further contribute to the development and maintenance of BDD. However, currently, the specific role that perfectionism plays remains theoretical due to a lack of appropriately designed empirical research. This issue also applies generally to all the developmental and maintenance factors for BDD, which remain



unknown due to this disorder being under-researched relative to other disorders of similar prevalence (Feusner, Neziroglu, Wilhelm, Mancusi & Bohon, 2010; Neziroglu & Barile, 2017).

Perfectionism is not determined by a diagnosis of BDD but does appear to be a common associate (Buhlmann, Etcoff & Wilhelm, 2008; Hartmann, Thomas, Greenberg, Matheny & Wilhelm, 2014). For example, 69% of individuals with a diagnosis of BDD were found to endorse the belief “I have to have perfection in my appearance” (Veale et al., 1996). Individuals with BDD can also hold variants of perfectionistic standards, where some may strive to achieve perfectionism or symmetry, whereas others may wish to achieve an ‘average’ appearance to allow them to blend in with others (Veale, 2004; Veale & Neziroglu, 2010). The empirical evidence demonstrates that adults with BDD are more likely to score higher on measures of perfectionism compared to healthy controls (Buhlmann et al., 2008; Hartmann et al., 2014; Schieber, Kollei, deZwaan, Müller, & Martin, 2013). This is understandable given the close relationship between BDD and Obsessive-Compulsive Disorder (OCD), where perfectionism is identified as one of the six key cognitive features. For this reason, measures of perfectionism were initially developed based on specific measures of OCD (Egan, Wade, & Shafran, 2011). In line with this concept, Blakey, Abramowitz, and Mahaffey (2016) investigated whether obsessive beliefs predict body image disturbance in a non-clinical sample of US undergraduate students (mean age = 20 years, 45% were male). After controlling for gender, OCD symptoms, general distress, and eating attitudes, perfectionism (assessed by a subscale on an obsessive belief measure) was the only type of obsessional belief found to significantly predict body image disturbance. The authors suggested that this is consistent with the theoretical models of BDD and makes logical sense, as holding maladaptive beliefs about the importance of body perfection will overlap with the typical obsessive preoccupations seen in individuals with BDD.

There is a limited literature on the association between perfectionism and appearance-related concerns in young people, where the available findings have typically been dissected from studies primarily conducted in the context of eating disorders. Subsequently, a broad range of outcome measures have been used, and not surprisingly, of the studies that have been published, mixed findings exist. Three separate studies using large school adolescent samples (aged between 12-16 years) found no significant correlations between perfectionism and measures of: appearance evaluation (Chen, Fox, Haase & Ku, 2010); perception of appearance (Custers & Van den Bulck, 2009); and body part satisfaction (Shaw, Stice & Springer, 2004). However, these studies all measured perfectionism using a subscale of the ‘Eating Disorders Inventory’, which is a six-item scale assessing perfectionism in the context of eating disorders

(Garner, 1991). Subsequently, utilising such a brief and contextual measure of perfectionism may somewhat explain these insignificant findings.

Dour and Theran (2011) found that in a group of 12-14-year olds in the US (46% male), maladaptive perfectionism (i.e. perceived discrepancy between achievement and high standards) significantly correlated with self-evaluation of appearance, but not with their perceived evaluation by others (termed “attribution”). Similarly, maladaptive perfectionism significantly predicted self-evaluation of appearance, which was comparable in boys and girls, whereas “attribution” was predicted to a smaller effect and was greater in boys compared to girls. Therefore, maladaptive perfectionism is more likely to drive the way we feel about our own appearance but has a lesser effect on how we perceive others to view our appearance. In addition, Australian pre-adolescent children (mean age = 9 years, 54% male) were studied to explore the predictors (including perfectionism) of muscle preoccupation (a subtype of BDD), as well as weight preoccupation (Saling, Ricciardelli, & McCabe, 2005). Using the cross-sectional data obtained, perfectionism was found to significantly predict muscle preoccupation whilst controlling for body mass index, school grade, self-esteem, parental and peer relations, and negative affect. As might be predicted, this association had a larger effect for boys compared to girls. Moreover, across a 10-month period, mean scores of muscle preoccupation decreased and perfectionism only remained a predictor for boy’s muscle preoccupation and not girls.

Perfectionism can be considered a multidimensional construct, and some tools are designed specifically to capture these variants. For instance, the Multidimensional Perfectionism Scale (MPS; Hewitt & Flett, 1991) consists of three dimensions, although only two are utilised for children/adolescents. That is: “Socially-Prescribed Perfectionism” (SPP; the belief that others hold high perfectionistic standards/expectations for the individual) and “Self-Oriented Perfectionism” (SOP; the belief that the individual holds perfectionistic standards for themselves). In adult university and community samples, evidence suggests that SPP is more strongly associated with appearance-related concerns compared to SOP (Bartsch, 2007; Hanstock & O’Mahony, 2002; Sherry et al., 2009). So rather than perfectionism been driven by a need to meet their own (unrealistic) personal standards, these findings indicate that the sample are trying to meet the standards set by others. Bartsch (2007) described this as being congruent with sociocultural theories, which argue that the rise in body image concerns in Western countries are a consequence of societal pressures placing more value on appearance. In contrast, Veale, Kinderman, Riley, and Lambrou (2003) demonstrated that adult patients with BDD displayed discrepancies between their perception of how they believe they look (“self-actual”), how they would ideally like to look (“self-ideal”), and how they feel they should look

("self-should"). Yet, there were no discrepancies found between their "self-actual" domain and both their perceptions of how others see them ("other-actual") and the perceived ideal of others ("other-ideal"). Therefore, the authors concluded that this clinical group are more concerned with a failure to achieve their own aesthetic standards, rather than the standards set by others. However, no measure of perfectionism was used in this study and so the findings cannot be directly comparable to those which use the MPS and demonstrate specific interpersonal perfectionism-related processes (such as SOP and SPP).

With regards to the relationship between appearance-related concerns and different dimensions of perfectionism in young people, only one study has utilised the child/adolescent version of the MPS but used it as one global measure and failed to discriminate between the dimensions. This was the study described earlier by Saling and colleagues (2005), who found that total perfectionism scores predicted muscle preoccupation in girls and boys. Two further studies included a sample with a mean age below 18 years and studied appearance-concerns in specific sporting populations (Dunn, Craft, Dunn & Gotwals, 2011; Ferrand, Magnan, Rouveix & Filaire, 2007). However, their samples were also inclusive of adults and they utilised the adult-version of the MPS, meaning that the findings were not uniquely applicable to adolescents. Nonetheless, they found that both SOP and SPP yielded significant correlations with appearance-related measures (Dunn et al., 2011), and that higher body esteem was associated with lower SPP but not SOP (Ferrand et al., 2007). Consequently, the relationship between perfectionism and appearance-related concern remains to be a significantly understudied area in children and adolescents.

Taken together, the literature is relatively sparse with regards to the role perfectionism plays in BDD and appearance-related concerns generally, and specifically with how this presents in young British people. Perfectionism has been demonstrated to be associated with several disorders in children; however, no reliable conclusion can be drawn regarding whether appearance-related concerns are driven by standards set by the self or others. Based on this, the current study would hypothesise that, overall, perfectionism would be positively associated with self-reported appearance-related concerns in a sample of British school children. Although assumptions about whether this is driven by SOP or SPP mechanisms cannot be made at this time.

### 1.3. The Current Study

The aims of the current study are two-fold; (1) to collect normative data on the extent that British adolescents are concerned about their appearance, and (2) to explore the role between perfectionism and appearance-related concerns.

The key objectives will be to:

- Identify the extent that adolescents are concerned about their appearance and whether these vary across age and gender, as well as what specific body areas of concern they have [Study One].
- Identify whether there is an association between perfectionism and appearance-related concerns in young people, and whether this varies across different dimensions of perfectionism [Study Two].

## 2. METHODS

---

# STUDY ONE

---

### 2.1. Design

A cross-sectional design was used to explore appearance-related concerns in young people, and how they might vary across age and gender. Therefore, appearance-related concern was considered the dependent variable, with age and gender as independent variables.

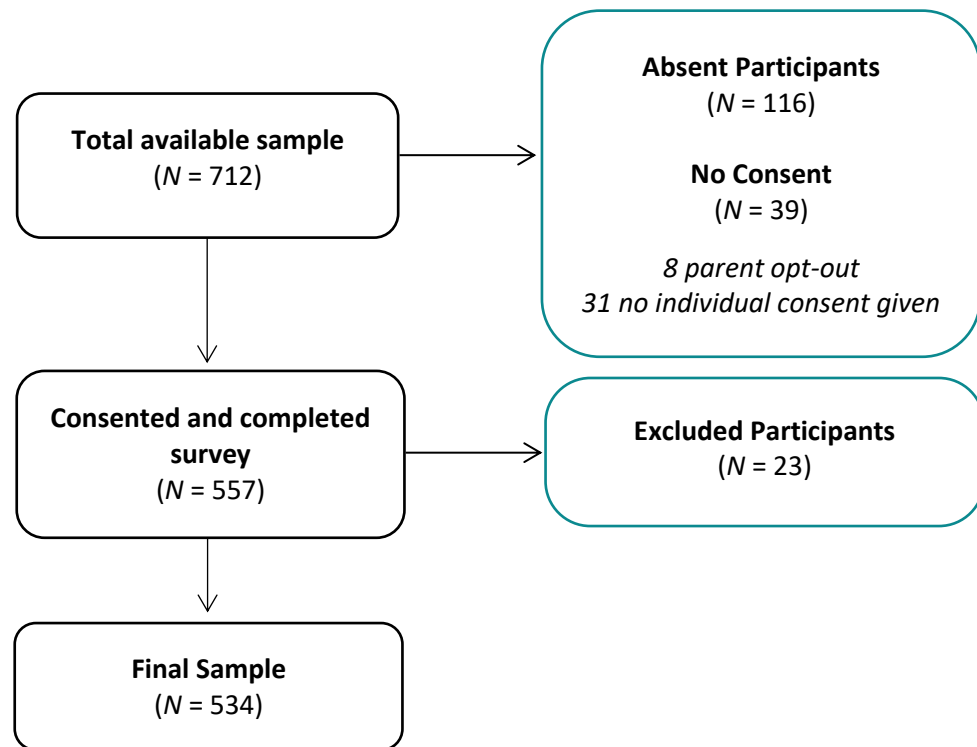
### 2.2. Sample

The sample consisted of secondary school students from a school located in a London borough. The inclusion criteria were any student studying in Years 7, 9, and 11 at the time of the research (aged 11-12, 13-14, and 15-16 years respectively). This was because of the exploratory nature of the study, where it was felt to be beneficial to obtain a sample that represented adolescents across every possible age within the secondary school. Utilising these three year groups achieved this, where the recruitment of every year group would have created unnecessary participant burden.

There were 238, 237, and 237 students enrolled in year 7, 9, and 11 respectively at the start of the academic year, indicating that 712 students were initially asked to participate in the study. Regarding the sample's ethnicity, the most predominant ethnicities were: White British (81.0%), White Other (3.9%), White and Black Caribbean (3.1%), Mixed Other (2.5%), and Black African (2.0%). Please see Appendix 6.1. for full ethnicity details. Exclusion criteria were students who: did not have parental consent to participate in the study; had more than two responses missing from one of the survey measures; and gave jovial and non-genuine responses (confirmed by a second rater).

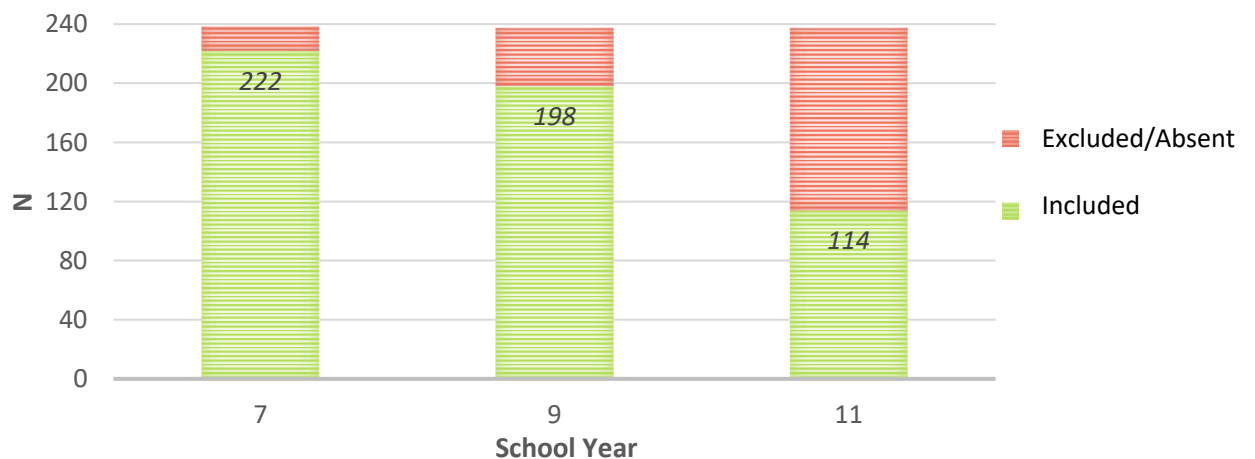
Figure 3 illustrates the process of obtaining the final sample. Eight students had parents who opted-out of the study, and 31 students recorded that they did not give their own consent to participate. In total, 557 students (293 male, 260 female, 4 gender unspecified) consented and completed the survey. A further 23 participants were excluded following an initial data screen (for missing/jovial responses), indicating that the final response rate for this study was 75.0% ( $N = 534$ ). Therefore, 116 students were absent for the administration of the survey (e.g. sickness,

study leave, exams, other educational activities) or declined to explicitly record their refusal to participate.



**Figure 3: Flow diagram to illustrate the recruitment of the final sample in Study One.**

Figure 4 below illustrates the final proportion of participants who were included in the research from each year group, of which 41.6% ( $N = 222$ ) of the sample were represented by Year 7, 37.1% ( $N = 198$ ) by Year 9, and 21.3% ( $N = 114$ ) Year 11. There were considerably fewer Year 11 participants compared to the other Year groups. Possible explanations for this are explored in the discussion.



**Figure 4: Bar graph showing the proportion of included participants in each school year.**

## 2.3. Measures

### 2.3.1. Body Image Questionnaire – Child and Adolescent version (BIQ-C)

The BIQ-C (Veale, 2009) was used as a measure of appearance-related concerns. It is a 12-item measure of BDD symptoms, where respondents are asked to score each item on a scale from zero (impaired) to eight (most impaired) – three of which are reversed (see Appendix 6.2.). Total scores range from 0 to 96, with higher scores reflecting greater impairment and a likely diagnosis of BDD. The questionnaire also allows for respondents to indicate the body features they are most concerned with and estimate how much time they spend thinking about these. The BIQ-C is used regularly as an outcome measure of BDD within the clinics associated with King's College London (KCL), and the wording was adapted from the adult version of the measure (BIQ; Veale, 2009). The BIQ-C has been found to have good internal consistency, is concurrent with the 'Body Dysmorphic Disorder Questionnaire – Adolescent Version' (Phillips, 2005), and can discriminate between groups of adolescents with "probable" BDD, "subthreshold" BDD, and non-BDD (Schneider et al., 2016<sub>a</sub>; 2016<sub>b</sub>; 2017<sub>a</sub>). Following the start of the current project, Schneider and colleagues (2016<sub>b</sub>) validated the BIQ-C in an Australian community sample of adolescents ( $N = 3,057$ ) and noted that it is an appropriate and brief measure for children over the age of 7, although they did find a 9-item version to be more adequate (BIQ-C-9). Subsequently they only note clinical cut-off scores for the BIQ-C-9, rather than the original BIQ-C, which was 41 in females and 25 for males (on the BIQ, the cut off is 59 for adults; Veale, 2009). Internal consistency of this scale in the current sample was high (*Cronbach's*  $\alpha = .90$ ).

# STUDY TWO

## 2.4. Design

A cross-sectional design was used to examine the relationship between appearance-related concerns and perfectionism. Appearance-related concerns were again considered the dependent variable, with perfectionism being the primary independent variable of interest. Anxiety, depression, and gender were considered as covariate variables.

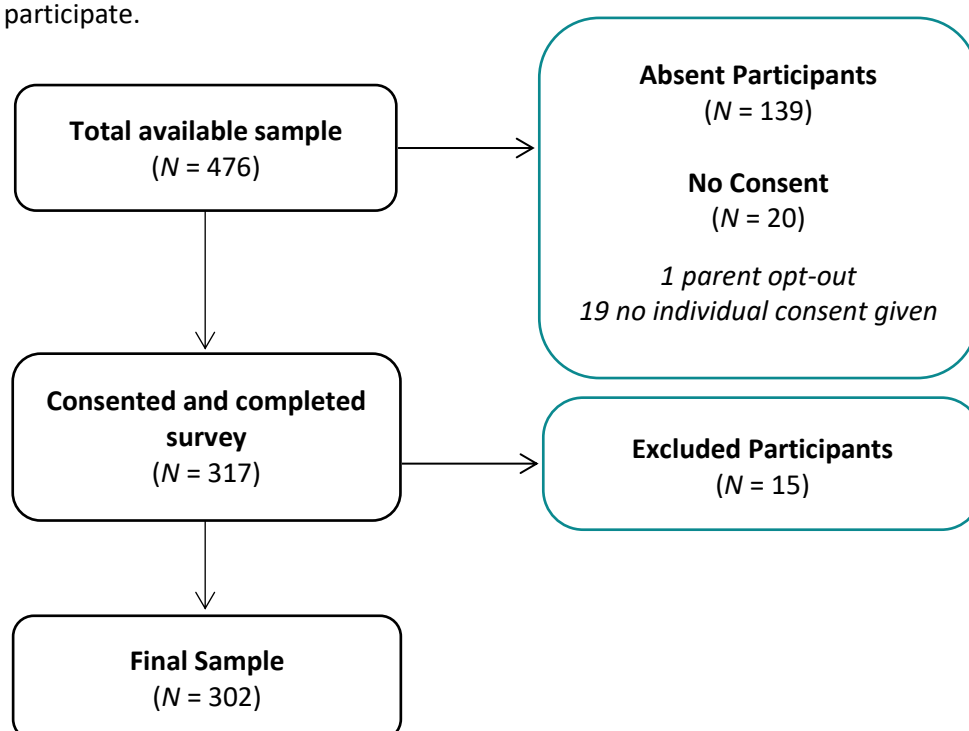
## 2.5. Sample

Secondary school students from the same school as Study One were recruited. Inclusion criteria were any student studying in Year 10 and 11 at the time of the research (aged 14-16). Of note,

the Year 11 sample is the same as that reported in Study One. Only pupils in Year 10 and 11 were recruited because the literature suggests that BDD symptoms are more prevalent in older adolescents compared to younger adolescents (Schneider et al., 2016<sub>a</sub>) and a larger sample size was not required for this study (please see ‘2.10. Power Analysis’ section).

There were 239 and 237 students enrolled in year 10 and 11 respectively at the start of the academic year, indicating that 476 students were initially asked to participate in the study. Regarding the sample’s ethnicity, the most predominant ethnicities were: White British (83.5%), White Other (3.6%), White and Black Caribbean (2.7%), Mixed Other (1.9%), and Black African (1.9%). Please see Appendix 6.3. for full ethnicity details. Exclusion criteria were students who: did not have parental consent to participate in the study; had more than two responses missing from one of the survey measures; and gave jovial and non-genuine responses (confirmed by a second rater).

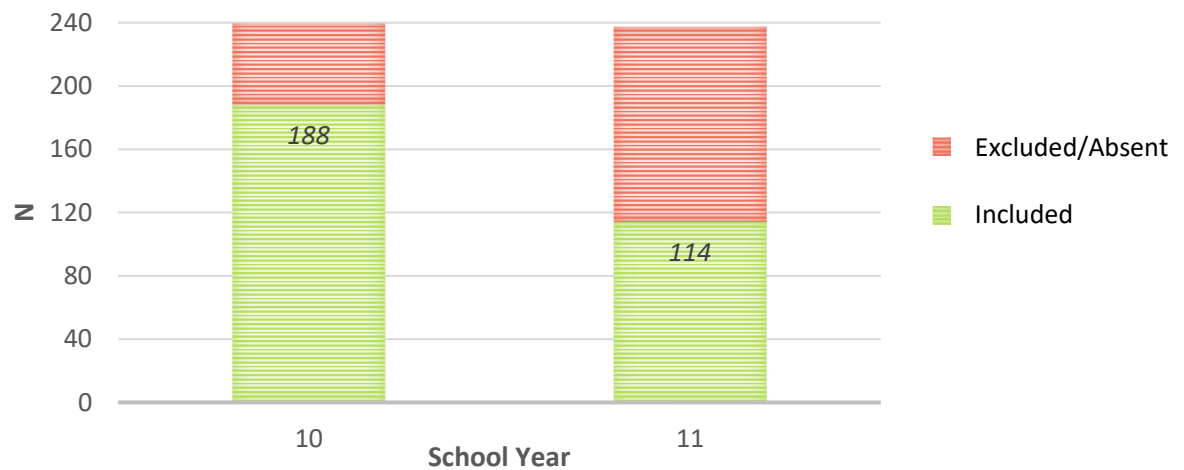
Figure 5 illustrates the process of obtaining the final sample. One of the students had parents who opted-out of the study, and 19 students recorded that they did not give their own consent to participate. In total, 317 students (150 male, 161 female, 6 gender unspecified) consented and completed the survey. A further 15 participants were excluded following an initial data screen (missing/jovial responses), indicating that the final response rate for this study was 63.4% ( $N = 302$ ). Therefore, 139 students were absent for the administration of the survey (e.g. sickness, study leave, exams, other educational activities) or declined to explicitly record their refusal to participate.



**Figure 5: Flow diagram to illustrate the recruitment of the final sample in Study Two.**



Figure 6 below illustrates the final proportion of participants who were included in the research from each year group, of which 62.3% ( $N = 188$ ) were represented by Year 10 and 37.7% ( $N = 114$ ) by Year 11.



**Figure 6: Bar graph showing the proportion of included participants in each school year.**

## 2.6. Measures

### 2.6.1. BIQ-C

As describe above in Study One, the BIQ-C was used as a measure of appearance-related concern. Reliability for this measure remained the same ( $\alpha = .90$ ).

### 2.6.2. Child-Adolescent Perfectionism Scale (CAPS)

The CAPS (Flett et al., 2016) is a 22-item self-report measure of perfectionism and is based on the Multidimensional Perfectionism Scale (MPS) developed by Hewitt and Flett (1991). The CAPS measures perfectionism across the two dimensions introduced earlier, Self-Oriented Perfectionism (SOP) and Socially-Prescribed Perfectionism (SPP). Participants are required to rate each item per a 5-point likert scale, ranging from one (“false – not at all true of me”) to five (“very true of me”). Higher scores represent greater perfectionism traits (see Appendix 6.4.). The psychometric properties of the CAPS have also been found to hold an adequate level of reliability and validity, and it was concurrent with another measure of perfectionism in this population (Flett et al., 2016). Cronbach’s alpha in the current study was .90 ( $\alpha = .84$  for SOP subscale;  $\alpha = .88$  for SPP subscale).

Of note, a recent systematic review of the validity and reliability of the tools used to measure perfectionism in children (<15 years) has been conducted (Leone & Wade, 2018), which includes the CAPS. The review highlights that the CAPS is the most widely used measure of perfectionism in children, and therefore is a good choice of measure given the vast amount of comparative data and robust psychometric properties. The authors further comment on the discrepancy in the factor structure found between various studies, where some publications use a three-factor structure, with SOP being split to represent “SOP-striving” and “SOP-critical”. Leone and Wade (2018) note that the authors of the CAPS criticised the original study that found that a three-factor structure was more adequate (McCreary, Joiner, Schmidt, & Ialongo, 2004), as they had used a non-standardised version of the CAPS questionnaire in their study. Subsequent research publications using the two-factor structure also predominantly report acceptable to very good levels of internal consistency for the SOP, with only 10% of the reviewed studies evidencing low values (Leone & Wade, 2018). However, this discrepancy in factor structure cannot be ignored, even more so when the three-factor structure has been supported in further research, such as that by O’Conner, Dixon, and Rasmussen (2009) with 15-16-year olds. This research group comment that “given that the factor structure of the CAPS changes as a function of age and ethnic group, we would recommend that the original 22-item version be administered initially for groups that differ in composition from the present sample” (p. 442). Subsequently, given that the current study is the first to explore the relationship between perfectionism and appearance-related concerns in a British sample, it was deemed appropriate to use the original, two-factor version of the CAPS.

#### 2.6.3. Revised Child Anxiety and Depression Scale – Short Version (RCADS-25)

The RCADS-25 (Ebesutani et al., 2012) is a shortened version of the original RCADS measure, which was developed by Chorpita, Yim, Moffitt, Umemoto, and Francis (2000). The shorter version was designed to reduce client burden/administration time; be a useful repeated measure for clinical practice; and provide wide-scale screening for anxiety and depression. It consists of 25 items, with 15 assessing the presence of anxiety and 10 assessing the presence of depression – consistent with the DSM-IV (see Appendix 6.5.). The depression items are those original to the RCADS, however the anxiety items have been reduced from 37-items. This has resulted in the RCADS-25 having one broad anxiety scale, rather than five separate subscales that correspond to five different anxiety disorders (e.g. social anxiety, panic etc.). It is a self-report measure and the young person is asked to indicate the frequency that each item applies to them on a 4-point likert scale ranging from zero (“never”) to four (“always”). Higher scores

represent greater clinical symptoms. The RCADS-25 has been found to have acceptable validity, as it could discriminate between youths with and without anxiety and affective disorder diagnoses, and has been shown to have reliable internal consistency, test-retest stability, as well as good convergent and divergent validity (Ebesutani et al., 2012; Muris, Meesters & Schouten, 2002). Cronbach's alpha for this study was .92 ( $\alpha = .87$  for depression subscale;  $\alpha = .87$  for anxiety subscale).

## STUDY ONE AND TWO

### 2.7. Procedure

Firstly, the school was approached and the study was explained to the Head Principal. Informed consent was subsequently obtained and a researcher-school agreement around the organisation of the study was negotiated. In return for the school's participation, a training session for teachers on appearance-related concerns was offered free of charge, and the researcher also offered to attend the Year 10 and 11 school assemblies to provide a talk on career pathways. Secondly, each student required parental consent to participate, and so two weeks prior to the survey administration date, a two-system method of obtaining 'opt-out' parental consent was initiated. This consisted of sending an information sheet and opt-out consent form: (1) directly to parents via 'parent mail' (the school's standard electronic method of communicating directly with parents), and (2) handing these documents to students via school letter to take home to their parents. This subsequently maximised the opportunities for parents to receive the information and opt-out if they wished. Parents were asked to sign and return the form to school, or contact the researcher on the email address provided, should they not wish for their child to participate in the research. Finally, before completing the survey and after being provided with an information sheet and survey instructions, eligible students were also asked to indicate their own consent to participate.

Between April and June 2017, surveys were administered during class time in a classroom setting (approximately 30 students per class), where each class had at least one survey facilitator (either the researcher, teacher, or A-Level psychology student volunteer). All facilitators were fully informed about the research and were given explicit instructions. It was organised that those students without parental consent went to the library with an alternative study task set by their teacher. Classes were either given a paper-copy of the survey or were provided with a web link

to access the survey electronically via the school computers. The instructions of the survey were delivered as a class, and each student was informed that their responses would remain confidential and that their participation was voluntary. They were further advised that their responses to the BIQ-C should not focus on weight/shape concerns, but strictly the appearance of body features. In return for their participation, students entered a raffle to win one of four £50 shopping vouchers. Due to the anonymity of the survey, students were required to create a “raffle-ID” made up of letters from their name, form class, and date of birth, so that the school could identify them should they be the raffle winner. In addition, students were informed that this ID would be used to identify them should they provide clinically concerning responses on the survey. In these cases, the raffle-ID codes would be shared with the school, who would decode and manage such individuals as per their standard pastoral procedures. Students were informed that this would always be managed privately between them and the school. All data was entered and stored in an electronic password-protected database.

## 2.8. Ethics

Full ethical approval was granted by the Psychiatry, Nursing, and Midwifery Research Ethics Subcommittee at KCL (see Appendix 6.6.).

## 2.9. Data Analytic Plan

All statistical analyses were carried out using IBM SPSS Statistics version 23. The data was initially screened for missing data and inappropriate responses. For each measure, if a participant’s response to an item was missing then it was replaced with the item average from the sample. Data was assessed for normality via visual inspection of histograms and Q-Q plots, and the skewness/kurtosis statistics were reviewed. Outliers were considered for all inferential statistical analysis.

With regards to Study One, descriptive statistics were primarily used to understand the prevalence and nature of appearance-related concerns. In addition, a 2x3 factorial analysis of variance (ANOVA) was used to understand how such concerns may differ according to gender (male/female) and age (school year 7/9/11), and whether these variables interacted. The statistical approach for Study Two began with descriptive statistics (including Pearson’s correlations) to understand the spread of the data and relationships between appearance-related concerns and dimensions of perfectionism, as well as with anxiety and depression. In addition, two multiple linear regressions were conducted to, firstly, identify whether

perfectionism generally is a unique predictor of appearance-related concerns, whilst controlling for gender, anxiety, and depression. And secondly, to explore whether different dimensions of perfectionism (SOP/SPP) have varying associations with appearance-related concerns, whilst also controlling for gender, anxiety, and depression.

The boundary of significance was held at  $p = .05$  for all analysis. This is the most commonly used level to make conclusions against the hypothesis being tested, suggesting that the probability the observed difference occurred by chance is less than 5% (Dorey, 2010).

## 2.10. Power Analysis

For Study One, a power analysis was not conducted due to the aims being exploratory in nature. Therefore, the authors aimed to survey as many students as possible. Regarding Study Two, the sample size required was calculated using the power calculator G\*Power (3.1.9.2). The total sample size (based on regression analysis) was calculated as 215 participants, which would have 80% power to detect a small-medium effect size at the 0.05 significance level. This effect size was based on the recent findings from a similar study conducted by Bernert and colleagues (2013). This looked at the relationship between eating disorder symptoms and perfectionism, whilst controlling for depressive symptoms. The  $t$  value from the reported regression analysis ( $\beta = .17$ ,  $t = 2.69$ ,  $p = <.01$ ) yielded a standardised effect size ( $r$ ) of 0.19. On this basis, a similar effect size was predicted to be found in this study. Considering the average 60% response rates obtained within academic studies (Baruch, 1999), it was approximated that at least 359 participants would need to be recruited.

## 3. RESULTS

---

### 3.1. Data Screening

The distribution of BIQ-C scores was reviewed and in doing so, it was apparent that the data was positively skewed. Arguably, this was expected due to this measure being designed for clinical use, whereby you could assume that (using a normative sample), you would obtain a larger proportion of respondents on the lower, non-clinical, end of the scale. Schneider and colleagues (2016<sub>a</sub>) also observed not normally distributed data in their study using the BIQ-C. Transformation of the data was explored; however, this did not adequately alter the data to fit a normal distribution curve. Consequently, analysis went ahead using the original data. Nonetheless, the statistical tests used were robust, where the potential effect of skewness was attenuated by the large sample sizes of the two studies and using the raw data would arguably retain interpretative value.

Of note, due to not wishing to specify their gender, two students in Study One and three students in Study Two were not included in any of the analysis where gender was used as a variable of interest (including calculations of means, SDs, and the ANOVA and regression analysis).

## STUDY ONE

### 3.2. Descriptive Statistics

#### 3.2.1. Prevalence and area of appearance-related concerns

All students were asked to rank up to three body features that they were most dissatisfied with and 78.8% ( $N = 421$ ) indicated that they were concerned by at least one body feature.

Out of a maximum total of 1,602 ranks (3 ranks per person), 21.6% ( $N = 346$ ) responses were missing, 18.8% ( $N = 301$ ) students responded with “I am happy with my appearance”, and 0.8% ( $N = 13$ ) indicated that they would “rather not say”. Overall, 942 ranks (58.8%) were provided, resulting in the identification of 52 different body features (Figure 7).

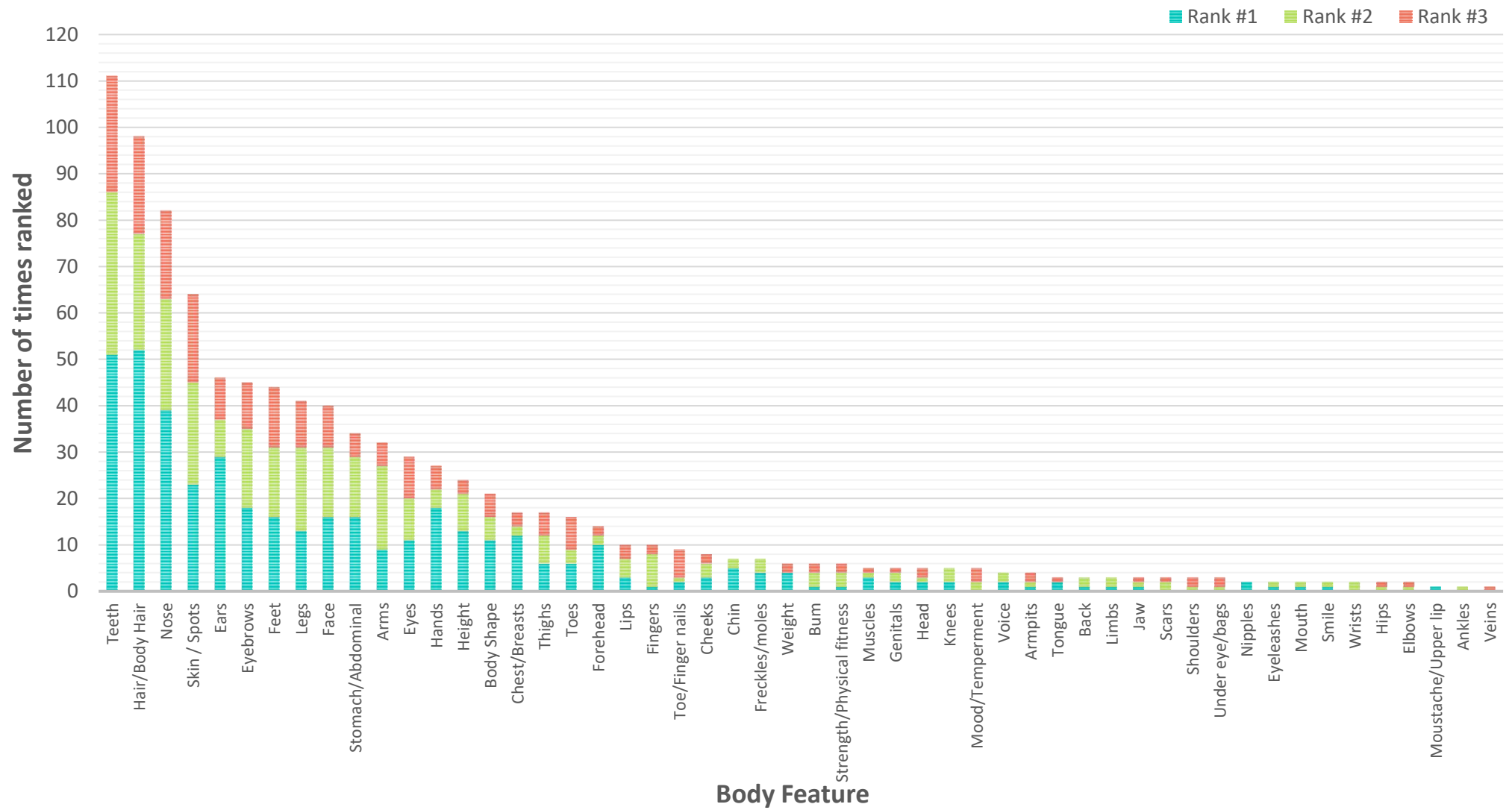


Figure 7: Bar graph showing the number of times each body feature was ranked by students.

Of the valid responses given ( $N = 942$ ), four body features were consistently ranked by the students as one of their three most concerning body features. That is: “teeth” (11.8%,  $N = 111$ ); “hair” or “body hair” (10.4%,  $N = 98$ ); “nose” (8.7%,  $N = 82$ ); and “skin” or “spots” (6.8%,  $N = 64$ ).

Students were asked to also indicate the reasons why they were concerned with the body features they had ranked. The typical responses for “teeth” were that the students did not like that they were: not straight; not white; too large or “fang” like; currently wearing braces; and had gaps. For “hair”, students often suggested that their hair was: not the right colour (ginger or too dark); not the right length; not the right thickness; too greasy/oily; there was too much of it; and difficult to style/manage. The most common reason for “nose” was that it was not the right shape (size, length, straightness, or the tip pointed upwards), or had imperfections specific to that area (spots or blackheads). Finally, dissatisfaction with the “skin” was most commonly associated with having “spots” or skin conditions (e.g. acne or eczema), not being the right colour, or being too hairy or dry.

### 3.2.2. Time spent on Appearance-related Concerns

As part of the BIQ-C, participants were asked to estimate how many hours and/or minutes they spend thinking about their appearance. The sample’s descriptive statistics are given in Table 9.

Due to the high variance in responses (as evident by large SDs), median and mode estimates were also provided. Females appear to spend considerably more time thinking about their body features of concern compared to males, although both male and females reported spending up to 1440 minutes (24 hours). When observing the year group means, Year 11s reported the least time on this activity; however, both the median and mode values suggested this group spend the longest. With regards to the males only, Year 11 boys were consistently observed to spend the longest thinking about their appearance-related concerns. For females, this is more variable across years, with each average estimate suggesting a different result.

Overall, there was huge variance in responses, as evident by the extremely large SDs, and therefore the reliability of these findings is questionable.



**Table 9: Descriptive statistics relating to the average length of time (minutes) students report spending thinking about their appearance-related concerns (*N* = 518).**

	<i>M</i>	<i>Median</i>	<i>Mode</i>	<i>SD</i>	<i>Range</i>
<b>TOTAL SAMPLE</b>	<b>84.71</b>	<b>10.00</b>	<b>0.00</b>	<b>237.618</b>	<b>0-1440</b>
Male	52.25	5.00	0.00	197.37	0-1440
Female	121.23	30.00	5.00	272.28	0-1440
<b>Year 7 Total</b>	<b>89.74</b>	<b>10.00</b>	<b>0.00</b>	<b>263.87</b>	<b>0-1440</b>
Male	45.24	5.00	0.00	189.36	0-1440
Female	141.29	20.00	5.00	323.26	0-1440
<b>Year 9 Total</b>	<b>81.86</b>	<b>10.00</b>	<b>0.00</b>	<b>206.66</b>	<b>0-1440</b>
Male	48.23	5.00	0.00	162.91	0-1439
Female	117.99	30.00	5.00	240.84	0-1440
<b>Year 11 Total</b>	<b>79.58</b>	<b>20.00</b>	<b>20.00</b>	<b>235.89</b>	<b>0-1440</b>
Male	74.93	10.00	10.00	264.50	0-1440
Female	86.09	30.00	20.00 <sup>a</sup>	207.63	0-1440

<sup>a</sup> Multiple modes exist, the smallest is shown.

### 3.2.3. Quality of Appearance-related Concerns

Table 10 presents the descriptive statistics and response frequencies for each of the 12-items on the BIQ-C. All items were rated on a scale of 0-8, with 0 representing responses of “not at all” or equivalent, and 8 represented responses equal or similar to “extremely” (each item had its own individualised responses, please see Appendix 6.2. for more detail).

On average, the highest mean scoring item on the BIQ-C was that referring to how “normal” or common the student’s appearance-related concern was, where students typically indicated that they believed only “some” people had similar concerns to them. The distribution of responses on this item was also one of the broadest, suggesting little agreement on any one response or response options grouped at one end of the scale. Similarly, the items regarding the importance of appearance and how unattractive their concerns were also saw variable responses, with a greater proportion of responses existing in the middle and top end of the scale compared to other items. Students indicated that their appearance-related concerns had minimal interference with their family relationships, school work, or social lives, and that they did not frequently avoid places or activities – evidenced by these items having the lowest averages.

Appearance-related concerns caused the most interference with romantic relationships, although still only to a small degree. Students most frequently indicated that they checked their appearance-related concerns up to five times a day; that they were “a little” noticeable to others, and that they were rarely-sometimes on their mind. Positively, students indicated very little distress associated with their appearance-related concerns.

**Table 10: Descriptive statistics and frequencies of responses for each BIQ-C item (*N* = 534).**

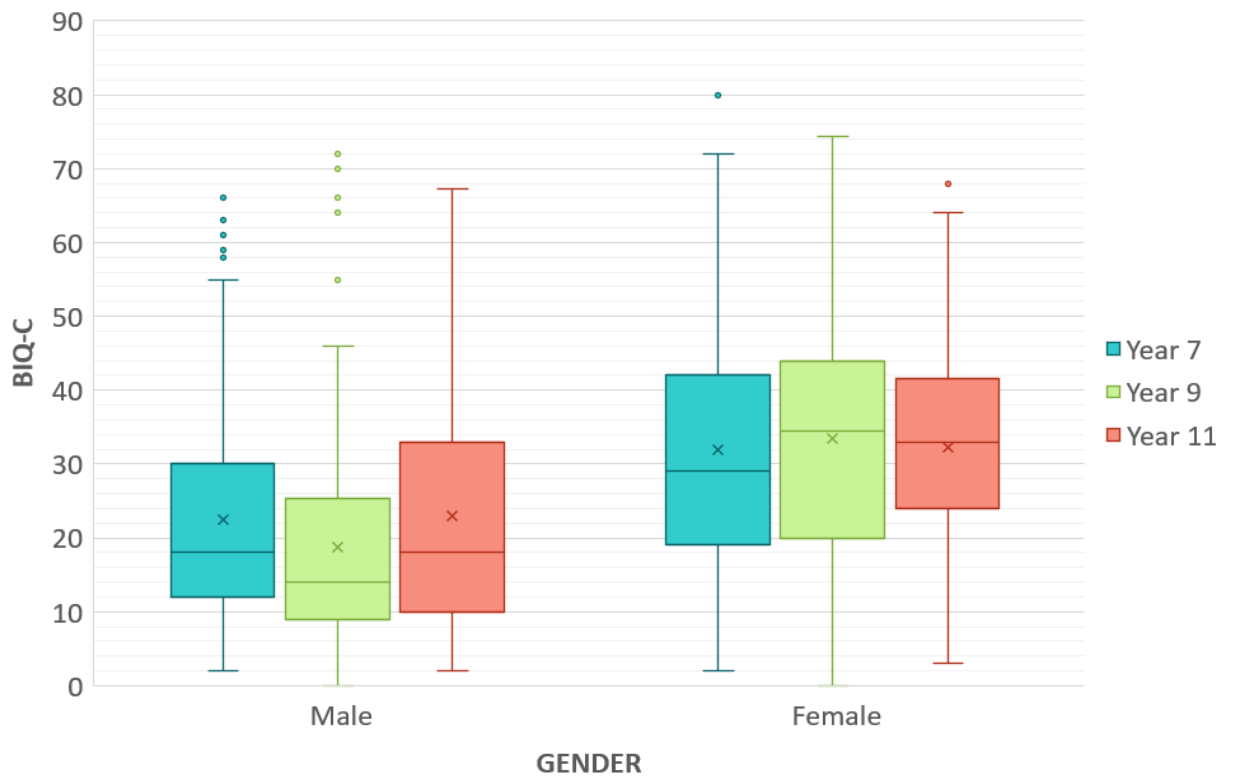
BIQ-C Item topic	<i>M</i> ( <i>SD</i> )	Frequency of response								
		Not at all 0	1	2	3	4	5	6	7	Extremely 8
Frequency of checking	2.59 (2.15)	14.6%	23.2%	<b>25.7%</b>	5.2%	11.4%	4.9%	8.8%	3.9%	2.2%
Perceived unattractiveness	2.99 (2.18)	13.3%	13.7%	<b>23.4%</b>	10.5%	16.5%	6.6%	9.4%	1.9%	4.9%
Distress caused	2.08 (2.12)	<b>31.5%</b>	15.2%	22.3%	6.0%	11.4%	3.7%	5.8%	1.5%	2.6%
Avoidance of places/activities	1.25 (1.87)	<b>54.1%</b>	14.0%	15.4%	2.8%	6.0%	2.2%	2.8%	1.1%	1.5%
Amount of preoccupation	2.84 (2.21)	17.2%	14.0%	<b>22.8%</b>	4.9%	21.0%	4.5%	9.2%	2.8%	3.6%
Interference with romantic relationships	2.01 (2.16)	<b>35.0%</b>	13.9%	20.3%	8.4%	9.0%	2.6%	6.2%	2.1%	2.6%
Interference with family relationships	0.82 (1.50)	<b>66.3%</b>	12.4%	11.0%	2.8%	3.0%	2.1%	1.5%	0.2%	0.7%
Interference with school work	1.04 (1.54)	<b>55.6%</b>	16.1%	15.0%	4.5%	3.6%	2.2%	2.6%	0.2%	0.2%
Interference with social life	1.26 (1.75)	<b>52.4%</b>	12.5%	16.3%	6.7%	6.2%	1.9%	2.1%	0.9%	0.9%
Importance of appearance	3.03 (2.28)	17.0%	9.7%	<b>20.6%</b>	13.7%	15.4%	4.9%	11.0%	2.4%	5.2%
Noticed by others	2.92 (2.14)	11.0%	<b>19.9%</b>	17.8%	17.8%	9.7%	10.9%	4.9%	4.1%	3.9%
How compares with others of same age, gender, ethnicity	3.80 (2.22)	8.1%	10.7%	11.4%	16.1%	12.5%	<b>17.0%</b>	9.9%	10.9%	3.4%

Frequencies in **bold** = most common response to item.

### 3.3. Gender and Age Differences in Appearance-related Concerns

Overall, the sample obtained a mean BIQ-C score of 26.62 ( $SD = 16.53$ ). Female students ( $M = 32.61$ ,  $SD = 16.23$ ) expressed greater appearance-related concerns compared to males ( $M = 21.20$ ,  $SD = 14.93$ ). Figure 8 presents the separate box plots for each gender and age group, where mean scores are represented by 'X'. Females consistently scored higher than males at each age on the BIQ-C. The gender difference was the greatest between male and female students of Year 9. There was also greater variability in the female's responses, and more male outliers, suggesting that more males significantly deviate from the "norm".

Of note, the total sample mean, and male and female mean scores, all fell below the clinical cut-off score that Veale (2009) suggested for adults (cut-off = 59). Moreover, 5.62% ( $N = 30$ ) of the total sample obtained a BIQ-C score of 59 or above.



**Figure 8: Box graph showing the BIQ-C descriptive statistics according to year group and gender.**

A 2x3 factorial ANOVA was conducted to determine whether the observed differences in BIQ-C score were significantly different. That is, between the two gender levels and three age levels, as well as exploring whether these variables interacted.

The ANOVA was initially run ( $N = 532$ ) and a statistically significant main effect of gender on BIQ-C score was identified ( $F(1, 526) = 63.037, p < .001, \text{partial } \eta^2 = .107$ ), whereby females had significantly higher scores on the BIQ-C compared to males. There was no main effect of age ( $F(2, 526) = .406, p = .666, \text{partial } \eta^2 = .002$ ) on BIQ-C scores, and no statistically significant interaction ( $F(2, 526) = 1.761, p = .173, \text{partial } \eta^2 = .007$ ). There was homogeneity of variances, as assessed by Levene's test for equality of variances ( $p = .234$ ), however, 19 outliers were detected via visual inspection of a box plot containing the residuals of the fitted model. Therefore, the ANOVA was run a second time after removing these outliers ( $N = 513$ ).

On this occasion, the assumption of homogeneity of variances was violated ( $p < .001$ ). Although, again, the ANOVA showed a statistically significant main effect of gender on BIQ-C score ( $F(1, 507) = 87.337, p < .001, \text{partial } \eta^2 = .147$ ), and no main effect of age ( $F(2, 507) = 1.314, p = .270, \text{partial } \eta^2 = .005$ ). Contrastingly, a statistically significant interaction between gender and year group for BIQ-C score was observed ( $F(2, 507) = 3.257, p = .039, \text{partial } \eta^2 = .013$ ), where simple main effects analysis revealed that males in Year 9 had BIQ-C scores that were statistically significantly lower compared to males of Year 11. The difference between males in Year 9 and 7 was just shy of significance ( $p = .058$ ). Of note, separate supplementary analysis accounting for the violation of the Levene's test all remained supportive of the findings (including an independent T-Test, one-way Welch ANOVA, and non-parametric tests – see Appendix 6.7.). That is, that there was a statistically significant main effect of gender, but not for age.

Taken together, it can be argued that BIQ-C scores are significantly different between male and female students, irrespective of variance and outliers, however, the observation that gender and year group interact upon BIQ-C scores is more ambivalent and should be interpreted with caution.

## STUDY TWO

### 3.4. Descriptive Statistics

The means, SDs, and ranges of the sample's appearance-related concerns, anxiety, depression, and perfectionism scores are presented in Table 11. On average, the total sample did not express appearance-related concerns in the clinical range (as suggested by Veale, 2009). When comparing males and females on these scales however, it is evident that females experienced more clinical symptoms of BDD, anxiety, and depression compared to males. This gender

difference was found to be statistically significant for BIQ-C ( $t(297) = -6.041, p < .001$ ), anxiety ( $t(297) = -5.030, p < .001$ ), and depression scores ( $t(297) = -2.368, p = .019$ ). Total perfectionism scores were comparable between the two genders, where females were observed to express greater SOP compared to males, and males expressed higher SPP – these differences were not statistically significant.

**Table 11: Means, standard deviations, and ranges of study variables ( $N = 302$ ).**

	Total Sample		Males		Females	
	<i>M (SD)</i>	<i>Range</i>	<i>M (SD)</i>	<i>Range</i>	<i>M (SD)</i>	<i>Range</i>
BIQ-C	28.58 (17.77)	0-83	22.34 (17.24)	0-83	34.14 (16.51)	3-74
RCADS-25 Anxiety	9.25 (7.10)	0-38	7.15 (6.75)	0-38	11.13 (6.92)	0-31
RCADS-25 Depression	7.39 (5.40)	0-24	6.59 (5.51)	0-24	8.07 (5.26)	0-24
CAPS Total	61.64 (16.7)	22-106	61.62 (15.67)	32-106	61.91 (16.37)	22-103
CAPS SOP	35.44 (9.10)	12-60	34.97 (8.83)	12-56	36.02 (9.27)	12-60
CAPS SPP	26.20 (9.09)	10-50	26.65 (9.13)	10-50	25.89 (9.06)	10-47

#### 3.4.1. Correlations

All correlations between the study variables were statistically significant (Table 12). The BIQ-C was correlated with all perfectionism dimensions to a moderate effect. A slightly greater correlation was present between BIQ-C and the SPP dimension of perfectionism compared to SOP. Both anxiety and depression scales were also correlated with BIQ-C to a large effect.

**Table 12: Pearson's bivariate correlation ( $r$ ) matrix between study variables ( $N = 302$ ).**

	BIQ-C	RCADS-25 Anxiety	RCADS-25 Depression	CAPS Total	CAPS SOP	CAPS SPP
BIQ-C	-					
RCADS-25 Anxiety	.636**	-				
RCADS-25 Depression	.538**	.714**	-			
CAPS Total	.380**	.410**	.366**	-		
CAPS SOP	.329**	.350**	.248**	.884**	-	
CAPS SPP	.343**	.376**	.398**	.884**	.562**	-

\*\* . Correlation is significant at the .01 level (2-tailed).

### 3.5. Regression Analysis

#### 3.5.1. Total Perfectionism Score

A multiple linear regression was conducted to test whether perfectionism was predictive of the variance in appearance-related concern, whilst controlling for anxiety, depression, and gender. Six outliers were removed due to their standardised residual being greater than  $\pm 3$  SDs, as indicated by the “casewise diagnostics” output. There was linearity as assessed by partial regression plots and a plot of studentized residuals against the predicted values. There was independence of residuals, as assessed by a Durbin-Watson statistic of 1.727. There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. There was no evidence of multicollinearity, as assessed by tolerance values greater than 0.1. There were no leverage values greater than 0.2 and values for Cook's distance above 1. The assumption of normality was met, as assessed by Q-Q Plot.

Using the enter method, the regression model was found to be statistically significant ( $F(4, 288) = 83.199, p < .001$ ), with perfectionism accounting for 53% of the variance in appearance-related concern ( $R^2 = .536, R^2_{adjusted} = .530$ ). As demonstrated in Table 13, perfectionism was found to be a unique predictor of the variance in BIQ-C score, independent of symptoms of anxiety and depression, and gender. Anxiety, depression, and gender were also found to be unique predictors of the variance in BIQ-C.

**Table 13: Statistics from the multiple linear regression predicting BIQ-C scores ( $N = 293$ ).**

	<i>B</i>	<i>SE<sub>B</sub></i>	$\beta$	<i>t</i>	<i>p</i>	<i>pr</i>
RCADS-25 Anxiety	1.020	.150	.422	6.804	< .001	.372
RCADS-25 Depression	.533	.187	.167	2.852	.005	.166
Gender	7.383	1.459	.214	5.059	< .001	.286
CAPS Total	.197	.048	.182	4.085	< .001	.234

*B* = Unstandardized regression coefficient; *SE<sub>B</sub>* = Standard error of regression coefficient;  $\beta$  = Standardized regression coefficient; *t* = t-test statistic; *p* = significance value; *pr* = partial correlation.

Due to gender being a binary variable, the interpretation was more difficult compared to the other continuous variables. Subsequently, a hierarchical multiple regression was run to determine the value of the adding gender (step 2), and then total CAPS score (step 3), in the prediction of the variance in BIQ-C scores, over and above anxiety and depression scores alone (step 1). The addition of gender (step 2) led to a statistically significant increase in  $R^2$  of .035

( $\Delta F(1, 289) = 20.678, p < .001$ ). Additionally, due to the cross-sectional design of this study, the model was reversed to see if appearance-related concerns predict perfectionism, whilst controlling for anxiety, depression, and gender. This model was also found to be significant ( $F(4, 288) = 22.267, p < .001, R^2 = .236, R^2_{adjusted} = .226$ ).

### 3.5.2. SOP/SPP Dimensions of Perfectionism

A second multiple linear regression was conducted to determine whether different dimensions of perfectionism were better/less able to predict the variance in appearance-related concerns, whilst controlling for anxiety, depression, and gender. Six outliers were removed due to their standardised residual being greater than  $\pm 3$  SDs, as indicated by the “casewise diagnostics” output. There was linearity as assessed by partial regression plots and a plot of studentized residuals against the predicted values. There was independence of residuals, as assessed by a Durbin-Watson statistic of 1.735. There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. There was no evidence of multicollinearity, as assessed by tolerance values greater than 0.1. There were no leverage values greater than 0.2 and values for Cook's distance above 1. The assumption of normality was met, as assessed by Q-Q Plot.

Using the enter method, the regression model was found to be statistically significant ( $F(5, 287) = 66.329, p < .001$ ), and accounted for 53% of the variance in appearance-related concern ( $R^2 = .536, R^2_{adjusted} = .528$ ). Table 14 offers the statistical output of the regression. When exploring each predictor, SOP was found to be a unique predictor of the variance in BIQ-C, independent of SPP, anxiety, depression and gender. SPP was not found to be a significant predictor of the variance in BIQ-C. Anxiety, depression, and gender were also found to be unique predictors of the variance in BIQ-C.

Similarly, another hierarchical multiple regression was run to determine the value of the adding gender (step 2), and then SOP and SPP score (step 3), in the prediction of the variance in BIQ-C scores, over and above anxiety and depression scores alone (step 1). The addition of gender (step 2) led to a statistically significant increase in  $R^2$  of .035 ( $\Delta F(1, 289) = 20.678, p < .001$ ).

**Table 14: Statistics from the multiple linear regression predicting BIQ-C scores ( $N = 293$ ).**

	$B$	$SE_B$	$\beta$	$t$	$p$	$pr$
RCADS-25 Anxiety	1.020	.151	.422	6.753	< .001	.370
RCADS-25 Depression	.535	.192	.168	2.793	.006	.163
Gender	7.376	1.468	.214	5.024	< .001	.284
CAPS SOP	.201	.094	.106	2.136	.034	.125
CAPS SPP	.193	.100	.100	1.929	.055	.113

$B$  = Unstandardized regression coefficient;  $SE_B$  = Standard error of regression coefficient;  $\beta$  = Standardized regression coefficient;  $t$  = t-test statistic;  $p$  = significance value;  $pr$  = partial correlation.



## 4. DISCUSSION

---

The aims of the two studies were to (1) collect normative data on the extent that adolescents are concerned about their appearance, and (2) explore the relationship between perfectionism and appearance-related concerns – the first project to investigate these concepts using a British school sample.

### 4.1. Summary – Study One

Overall, appearance-related concerns were a particularly common phenomenon in this sample of adolescents, as almost 80% of the sample offered at least one body feature of concern. This is greater in comparison to the previous figure provided by Schneider and colleagues (2016<sub>b</sub>), which was 49.5% in Australian 12-18-year olds. This difference in finding may be due to the cultural variations in the samples used. Of interest, the current finding of 80% was the most comparable to the rate found in Lambrou and colleagues' study (2012). This study also used a sample from London, albeit they were adult participants, and found that 88% identified at least one body feature they disliked. In the current study, the most common body areas of concern were: teeth, hair/body hair, nose, skin/spots, which have also been noted in previous literature using both normative and clinical samples. For example, in a normative sample of 308 German 15-21-year olds (87% female), skin was ranked the most common feature of concern, followed by chest area and nose (Möllmann et al., 2017); in a sample of adolescents ( $N = 36$ , 14-20 years) and adults ( $N = 164$ , aged over 21 years) with current or past diagnoses of BDD, both skin and hair were consistently ranked as the greatest features of concern (Phillips et al., 2006); and Australian school children who indicated subthreshold or probable BDD symptoms ( $N = 162$ , aged 12-18 years, 57% male) ranked skin as the most prevalent area of concern (Schneider et al., 2017<sub>b</sub>). The main difference with the current sample was the implication of teeth being so highly ranked, as this feature was not implicated in the previous studies. This observation, that normative appearance-concerns do resemble those seen in patients with BDD, suggests that both normative and clinical appearance-related concerns share some qualitative properties, and therefore supports the proposition that they are part of the same spectrum (Lambrou et al., 2012; Phillips, 2005).

Regarding the characteristics of the expressed appearance-related concerns, overall the sample indicated very little distress, avoidance, or interference, which are core features of BDD. This is consistent with expectations given the normative sample studied and confirms that young

peoples' experiences of appearance-related concerns are not clinically significant but are quantitatively different from those seen in individuals with BDD. This was further supported by the observation that, on average, BIQ-C scores did not reach clinical threshold (as defined by the current literature on adults - Veale, 2009). In fact, on each of the BIQ-C items, no more than 5% of the sample reported concerns to the extreme end of the likert scale (i.e. point 8). This corresponds with the literature on prevalence rates, where approximately 1.7-3.6% of adolescents meet diagnostic criteria for BDD and would subsequently score in this extreme range (Mayville et al., 1999; Möllmann et al., 2017; Schneider et al., 2016<sub>a</sub>; Veale et al., 2016). Furthermore, most adolescents reported checking their appearance up to five times a day, although the findings on the amount of time specifically spent thinking about appearance-concerns was highly variable. This could imply that this finding was unreliable or could reflect a genuine high degree of variance. On the whole, the sample believed that appearance was somewhat an important aspect of their identity, and thus understandably expressed some preoccupation with their appearance. They also perceived their feature(s) to be a little unattractive and noticeable to others, and held the belief that, although their concerns were not uncommon, they were neither readily experienced by other people their age, gender, and ethnicity.

What was clear from the data was that girls consistently reported significantly higher appearance-related concerns compared to boys, across all age groups. This is concurrent with other literature (Mayville et al., 1999; Schneider et al., 2017<sub>a</sub>; Webb et al., 2015; Zimmer-Gembeck et al., 2017), suggesting that (in non-clinical samples) appearance-related concerns are more prevalent in girls compared to boys. The current findings did not, however, illustrate that appearance-related concerns differed across age as has been found in other research with young people (Rosenblum & Lewis, 1999; Schneider et al., 2016<sub>a</sub>; Zimmer-Gembeck et al., 2017). It is possible that this was a consequence of the smaller proportion of older adolescents in this sample. Similarly, there was some indication that an interaction between gender and age may be present, although this evidence was not strong enough to make any reliable conclusions and warrants further investigation.

Taken together, the findings from Study One have helpfully given an insight into the experience of appearance-related concern for British adolescents. Some of which are consistent with studies from other cultures/ethnicities, including the greater report of concerns in girls compared to boys, and the similarity in the features of concern identified. However, findings specific to this British sample and therefore novel to the literature include the implication of teeth as the most common area of concern and the specific characteristics of the concerns (e.g.

distress, checking, interference, etc.). This study has also provided initial estimations of the BIQ-C norms for British adolescents, although further research samples are required.

## 4.2. Summary – Study Two

Regarding the relationship between perfectionism and appearance-related concerns, as a total construct, perfectionism was significantly correlated with and accounted for a significant amount of variance in the prediction of appearance-related concerns, whilst controlling for gender, anxiety, and depression. As a model, these factors collectively accounted for over half the variance in adolescent appearance-related concerns and indicated a strong association. As discussed earlier, this is understandable given that perfectionistic beliefs are regularly featured within BDD psychological models, and how BDD is related to OCD (where the role of perfectionism is well-established). Nonetheless, of all the variables inputted into the regression model, perfectionism was the least influential compared to anxiety, depression, and gender. This suggests that, although perfectionism was associated with appearance-related concerns in the current sample, anxiety, depression, and gender are more likely to have an effect on this relationship. Therefore, it is important that these variables continue to be explored and measured in future studies and clinical practice.

It was only Self-Oriented Perfectionism (SOP), and not Socially-Prescribed Perfectionism (SPP), that accounted for unique variance in the prediction of appearance-related concerns (whilst controlling for gender, anxiety, and depression). This suggests that, although both dimensions were correlated with BIQ-C scores, perfectionistic standards in British adolescents are more likely to be driven by the self. That is, young people with appearance-related concerns are more likely to have a stringent evaluation of one's own behaviour, be avoidant of failures, and compulsively strive for perfection and self-improvement, rather than believing that they need to attain standards prescribed by significant others or extrinsic reward (Hewitt & Flett, 1991). This intrinsic motivation then, could explain why seeking reassurance from others is often not adequate in diminishing appearance-related distress, and why self-focussed attention and other cognitive biases are present within BDD formulations and models (Fang & Wilhelm, 2015; Veale, 2004; Veale & Neziroglu, 2010; Wilhelm & Neziroglu, 2002; Wilhelm et al., 2012).

This finding, that intrinsically driven perfectionism (SOP) predicts appearance-related concerns, conflicts with several previous studies which all noted SPP to be the most influential dimension of perfection. For instance, only SPP was found to yield a significant correlation with dysmorphic concerns in female Australian University students, and body image disturbance in Canadian

University students (53% male), whereas SOP did not (Hanstock & O'Mahony, 2002; Sherry et al., 2009); only SPP was found to account for a significant amount of variance in dysmorphic concerns (whilst controlling for acne associated quality of life, past levels of acne, general well-being, and general well-being by past acne level interaction), whereas SOP did not (Hanstock & O'Mahony, 2002); both SPP and SOP accounted for a significant amount of variance in body image disturbance (whilst controlling for gender, body mass index, reassurance seeking) and dysmorphic concern (whilst controlling for self-esteem, depression, and gender), although in both cases this relationship was greater in SPP compared to SOP (Bartsch, 2007; Sherry et al., 2009); and those indicated as having probable BDD scored significantly higher on SPP compared to participants not considered to have BDD, where their SOP scores did not significantly differ (Bartsch, 2007). However, it is important to note that these studies utilised Australian and Canadian adult university samples, and so a cultural and age difference may explain this discrepancy. Additionally, Sherry's research group (2009) consisted of university students who specifically belonged to a fitness centre, and therefore would not be representative of a normative community sample.

When investigating appearance-related concern in the absence of SOP and SPP dimensions of perfectionism, or even perfectionism altogether, there has been some support for the role of intrinsic mechanisms compared to those driven by societal factors. For instance, Dour and Theran (2011) found that maladaptive perfectionism was found to be associated with self-evaluations of appearance, and not other people's evaluations. Veale and colleagues (2003) also identified that adult BDD patients were more preoccupied with achieving their own aesthetic standards rather than those set by others. Further, factor analysis has demonstrated that SOP is more closely associated with positive affect, and SPP with negative affect (Cox, Enns & Clara, 2002). Subsequently it would make sense that SOP was found to hold a stronger relationship with appearance-related concerns in the current study, due to it being a non-clinical sample and capturing primarily normative concerns. Attributing one dimension of perfectionism to appearance-related concerns may also be too simplistic here, and account for the inconsistencies in findings. Instead, this relationship could be better explained through a transitional process, where SPP may transition into SOP over time, or vice versa. Subsequently, the categorisation of perfectionism dimensions may be insufficient should perfectionism have dynamic properties.

### 4.3. General Limitations

The findings of the current project should be considered alongside the study's limitations and interpreted with caution. Firstly, the cross-sectional design of the studies makes it difficult to understand the true association between the variables, for instance the temporal relationship and direction of effects between perfectionism and appearance-related concern in Study Two. Furthermore, the generalisability of these results is limited, where it cannot be assumed that they represent young people from other areas within or outside of Britain, or at other points in time. Within the current sample participants were not equally distributed among age groups, with there being considerably fewer Year 11 participants compared to other year groups. Although this is likely to be a consequence of the timing of the project (i.e. between the months of April and June which coincided with the GCSE examination period), it also suggests that there was a bias present between those Year 11 students who chose to respond during this time and those who did not. In addition, this study was designed to measure appearance-related concerns at one point in time, whereas a study by Rudiger, Cash, Roehrig, and Thompson (2007) demonstrated that body image is a variable construct and found that perfectionism was more strongly associated with this day-to-day variability compared to state levels.

Secondly, despite being a common method as it allows researchers to quantify psychological constructs from large samples with little participant burden, the questionnaires used in the current study were all reliant on self-report. This suggests that the findings may contain some error and/or response bias. For instance, responding in a social desirable manner, faking bad, or not feeling able to provide honest responses. The survey was administered in classroom setting, and therefore the conditions were not fully controlled. It is likely that some participants did not feel that this setting was private enough for them to answer honestly, especially as appearance-related concerns can be perceived a sensitive topic to some and is likely to promote classroom conversation. One parent did feedback to the researcher that, due to these conditions, her sons had wanted to communicate differently on their survey to what they did. Additionally, some participants may have not provided their full attention to the task, conferred with peers, or not fully understood the questions asked.

Screening the data revealed several jovial responses which were later excluded. Fan and colleagues (2006) found that these did not significantly bias a study's findings if large groups were used. Their evidence also indicated that adolescents in school settings are particularly likely to report extreme responses on self-report measures, indicating that the current findings may be an over/underestimation of young people's true appearance-related concerns. Obtaining corroborative evidence from parents or teachers, or conducting individual interviews,

may have eliminated some of this error, however this would have required significantly more time and resources, as well as sacrificing the sample size.

Finally, there was a lack of validation of the appearance-related concerns tool in British young people. As discussed earlier, since the start of this research Schneider and colleagues (2016<sub>b</sub>) validated the BIQ-C using an Australian sample and recommended that a 9-item version of the tool was more appropriate (BIQ-C-9). Yet clinically, the 12-item BIQ-C is considered adequate in measuring young people's appearance-related concerns in the UK, as it is routinely used in the national and specialist clinic. With it being a clinical tool, the data was skewed and therefore all statistics should be interpreted with some caution. Additionally, there is contrasting evidence about whether perfectionism and the two sub-scales are affected by age in young people (e.g. Arale, 2007; Castro et al., 2004), and whether a two- or three-factor structure is more adequate for the CAPS (Leone & Wade, 2018; McCreary et al., 2004; O'Conner et al., 2009). Alternatively, Sherry and colleagues (2009) argued that the role of perfectionism in the development of appearance-related concerns is not that the individual is striving to reach perfection, but rather that they are driven to avoid appearing imperfect to others (i.e. the expression of perfectionism in public). This group used a measure of perfectionism designed to explicitly measure this dimension and found that the non-display of imperfection was uniquely tied to body image disturbance beyond trait perfectionism (SOP and SPP), and other hypothesized contributors to this disturbance (gender, body mass index, and reassurance-seeking). Therefore, the use of alternative perfectionism outcome measures may have yielded different results.

#### 4.4. Implications & Future Directions

These findings were the first of their kind to attempt to understand appearance-related concerns in British young people. They provide guidance on the level of "normal" appearance concerns in young people, which could inform clinicians when assessing patients with BDD, by allowing them to compare the reported concerns to this baseline. Appearance-related advice for non-clinical groups can further be developed based on this normative information, working to educate and normalise such experiences. Even more, given the qualitative similarities observed between normative and clinical appearance concerns, this information could also potentially form early intervention work for BDD. This has already been conducted to target general body image problems in young people, that is, the "Happy Being Me" programme developed by Richardson and Paxton (2010). This is a theoretically derived school-based body image intervention (three 50-minute sessions) targeted for young adolescent girls and has

shown to improve risk factors for negative body dissatisfaction, including internalisation of the thin body ideal, body comparison, and appearance conversations at post-intervention and follow-up compared to a control group, as well as topic knowledge, body satisfaction, dietary restraint, and self-esteem (Richardson & Paxton, 2010). Moreover, it could be useful within clinical BDD treatment, again for educational purposes about the differences between “normal” and BDD concerns, but also as a guide for setting therapeutic goals (e.g. Study One demonstrated that 63.5% percent of young people estimated that they check their appearance five times a day or less, and therefore this would be a reasonable therapeutic target). Should further empirical evidence reliably demonstrate an implicating role of perfectionism, these findings could contribute towards the development of an optional module within psychological treatments, where if applicable, perfectionistic cognitive biases can be addressed and challenged. An appearance-related perfectionism outcome measure or an additional item on an already established tool may help decipher whether such module is required in treatment.

Regarding future directions, it is important to encourage other researchers to attempt to replicate the current study using other British samples to demonstrate the reliability of such findings. The current study was only able to capture appearance-related concerns and perfectionism at one point in time. Therefore, it may be useful to collect longitudinal data to explore how appearance-related concerns and perfectionism change over a significant period. This would also highlight how resilient these concerns are, indicating whether these constructs are better conceptualised as states or traits. Moreover, longitudinal research would provide insights into whether reported appearance-related concerns naturally resolve by adulthood or continue to develop to be clinically significant. It would also be helpful to include a British clinical group of patients with BDD, so explicit comparisons can be made, as well as providing an apt opportunity to more adequately validate the BIQ-C. Veale (2004) further advised that any study investigating risk factors of BDD should include controls from non-clinical samples, as well as those with OCD and depression, due to the similarity in phenomenology and reported comorbidity.

Moreover, it would be beneficial to incorporate several alternative study variables within the design, such as ethnicity and culture. As discussed, previous literature has demonstrated the effect ethnicity and culture have on the prevalence and manifestation of appearance-related concerns (Bartsch, 2007; Bohne et al., 2002; Boroughs et al., 2010; Cash et al., 2004; Marques et al., 2011; Mayville et al., 1999). Unfortunately, ethical regulations prevented the current study from specifically assessing ethnic differences in appearance-related concerns within this British sample, but this would be an important variable to explore and control for in future studies.

Other variables that have been shown to effect appearance-related concerns and so should also be covered within future research include: body mass index (Cash et al., 2004; Chen et al., 2010), self-esteem (Bartsch, 2007), self-compassion (Barnett & Sharp, 2016), aesthetic sensitivity (Schieber et al., 2013) and perceived teasing (Mastro, Zimmer-Gembeck, Webb, Farrell, & Waters, 2016).

Tylka (2012), who embodies the positive psychology perspective, highlighted the misassumption that low appearance-related disturbance is equivalent to positive body image, and so within the literature we cannot conclude that findings on negative body image is the opposite for positive body image (e.g. that greater perfectionism is associated with BDD symptomology, and so lower perfectionistic traits would result in reduced symptoms and a positive body image). Consequently, it may be worthwhile for future research to consider the role of positive body image and the factors associated with positive perceptions of appearance, rather than solely focusing on appearance concerns, disturbance, and anxieties. This could more accurately inform clinical treatment and the skills required to overcome such appearance-related disorders.

Subsequently, there are some valuable clinical implications from this paper, namely the inclusion of this data in the assessment and treatment of BDD, but also in the education of non-clinical groups and professionals. Additionally, to significantly contribute to this research area and enhance our understanding, future research should have a carefully considered design, and concurrently explore positive body image.

## 4.5. Conclusion

The current study has collected data on the frequency and presentation of appearance-related concerns in young British people, a warranted research area given that appearance-related concerns have been shown to vary according to ethnicity. Appearance-related concerns were found to be particularly common in young people, and these were associated with little distress and interference, unlike the experience of appearance-related concerns in BDD. The current findings also showed that perfectionism accounts for unique variance in the prediction of appearance-related concerns, over and above gender, anxiety, and depressive symptoms. More specifically, Self-Oriented Perfectionism (SOP) was found to significantly predict the variance in appearance-related concerns, whereas Socially-Prescribed Perfectionism (SPP) did not. Therefore, perfectionism appears to play an important role in the experience of appearance-related concern, although it is not clear whether it acts as a developmental or maintaining factor.



As such, it should continue to be included in psychological models and treatment of BDD, and further research is warranted to replicate and advance our understanding in this area.

## 5. REFERENCES

---

- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: American Psychiatric Association.
- Arale, K. M. (2007). *The connection between perfectionism and depression, anxiety, and anger with consideration of ethnicity* (Doctoral Thesis). Retrieved from ProQuest database (304818678).
- Barnett, M. D. & Sharp, K. J. (2016). Maladaptive perfectionism, body image satisfaction, and distorted eating behaviours among U.S. college women: The mediating role of self-compassion. *Personality and Individual Differences*, 99, 225-234.
- Bartsch, D. (2007). Prevalence of body dysmorphic disorder symptoms and associated clinical features among Australian university students. *Clinical Psychologist*, 11, 16-23.
- Baruch, Y. (1999). Response Rate in Academic Studies-A Comparative Analysis. *Human Relations*, 52, 421-438.
- Bento, C., Pereira, A. T., Maia, B., Marques, M., Soares, M. J., Bos, S., Valente, J., Gomes, A., Azevedo, M. H. P., & Macedo, A. (2010). Perfectionism and eating behaviour in Portuguese adolescents. *European Eating Disorders Review*, 18, 328-337.
- Bernert, R. A., Timpano, K. R., Peterson, C. B., Crow, S. J., Bardone-Cone, A. M., le Grange, D., Klein, M., Crosby, R. D., Mitchell, J. E., Wonderlich, S. A., & Joiner, T. E. (2013). Eating disorder and obsessive-compulsive symptoms in a sample of bulimic women: Perfectionism as a mediating factor. *Personality and Individual Differences*, 54, 231-235.
- Blakey, S. M., Abramowitz, J. S., & Mahaffey, B. L. (2016). Do obsessive beliefs predict body image disturbance? *Journal of obsessive-compulsive and related disorders*, 11, 96-100.
- Bohne, A., Keuthen, N. J., Wilhelm, S., Deckersbach, T., & Jenike, M. A. (2002). Prevalence of symptoms of Body Dysmorphic Disorder and its correlates: A cross-cultural comparison. *Psychosomatics*, 43, 486-490.

- Boroughs, M. S., Krawczyk, R., & Thompson, J. K. (2010). Body Dysmorphic Disorder among Diverse Racial/Ethnic and Sexual Orientation Groups: Prevalence Estimates and Associated Factors. *Sex Roles, 63*, 725–737.
- Buhlmann, U., Etcoff, N. L., & Wilhelm, S. (2008). Facial attractiveness ratings and perfectionism in body dysmorphic disorder and obsessive-compulsive disorder. *Journal of Anxiety Disorders, 22*, 540–547.
- Cash, T. F., Phillips, K. A., Santos, M. T., & Hrabosky, J. I. (2004). Measuring “negative body image”: Validation of the Body Image Disturbance Questionnaire in a nonclinical population. *Body Image, 1*, 363-372.
- Castro, J., Gila, A., Gual, P., Lahortiga, F., Saura, B., & Toro, J. (2004). Perfectionism dimensions in children and adolescents with anorexia nervosa. *Journal of Adolescent Health, 35*, 392-398.
- Castro-Fornieles, J., Gual, P., Lahortiga, F., Gila, A., Casula, V., Fuhrmann, C., Imirizaldu, M., Saura, B., Martinez, E., & Toro, J. (2007). Self-oriented perfectionism in eating disorders. *International Journal of Eating Disorders, 40*, 562-568.
- Chen, L., Fox, K. R., Haase, A. M., & Ku, P. (2010). Correlates of body dissatisfaction among Taiwanese adolescents. *Asia Pacific Journal of Clinical Nutrition, 19*, 172-179.
- Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L. A., & Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A revised child anxiety and depression scale. *Behavior Research and Therapy, 38*, 835–855.
- Cox, B. J., Enns, M. W., & Clara, I. P. (2002). The multidimensional structure of perfectionism in clinically distressed and college student samples. *Psychological Assessment, 14*, 365-373.
- Custers, K., & Van den Bulck, J. (2009). Viewership of Pro-Anorexia Websites in Seventh, Ninth and Eleventh Graders. *European Eating Disorders Review, 17*, 214–219.
- Dorey, F. (2010). In Brief: The P Value: What Is It and What Does It Tell You? *Clinical Orthopaedics and Related Research, 468*, 2297–2298.

- Dour, H. J., & Theran, S. A. (2011). The interaction between the superhero ideal and maladaptive perfectionism as predictors of unhealthy eating attitudes and body esteem. *Body Image, 8*, 93–96.
- Dunn, J. G. H., Craft, J. M., Dunn, J. C., & Gotwals, J. K. (2011). Comparing a domain-specific and global measure of perfectionism in competitive female figure skaters. *Journal of Sport Behavior, 34*, 25-46.
- Ebesutani, C., Reise, S. P., Chorpita, B. F., Ale, C., Regan, J., Young, J., Higa-McMillan, C., & Weisz, J. R. (2012). The Revised Child Anxiety and Depression Scale-Short Version: Scale Reduction via Exploratory Bifactor Modeling of the Broad Anxiety Factor. *Psychological Assessment, 24*, 833-845.
- Egan, S. J., Wade, T. D., & Shafran, R. (2011). Perfectionism as a Transdiagnostic Process: A Clinical Review. *Clinical Psychology Review, 31*, 203-212.
- Fan, X., Miller, B. C., Park, K., Winward, B. W., Christensen, M., Grotevant, H. D., & Tai, R. H. (2006). An exploratory study about inaccuracy and invalidity in adolescent self-report surveys. *Field Methods, 18*, 223–244.
- Fang, A., & Wilhelm, S. (2015). Clinical Features, Cognitive Biases, and Treatment of Body Dysmorphic Disorder. *Annual Review of Clinical Psychology, 11*, 187-212.
- Fang, A., Matheny, N. L., & Wilhelm, S. (2014). Body Dysmorphic Disorder. *Psychiatric Clinics of North America, 37*, 287-300.
- Ferrand, C., Magnan, C., Rouveix, M., & Filaire, E. (2007). Disordered eating, perfectionism and body-esteem of elite synchronized swimmers. *European Journal of Sport Science, 7*, 223-230.
- Feusner, J. D., Neziroglu, F., Wilhelm, S., Mancusi, L., & Bohon, C. (2010). What causes BDD: research findings and a proposed model. *Psychiatric Annals, 40*, 349–355.
- Flett, G. L., Druckman, T., Hewitt, P. L., & Wekerle, C. (2012). Perfectionism, coping, social support, and depression in maltreated adolescents. *Journal of Rational-Emotive & Cognitive-Behavior Therapy, 30*, 118-131.

- Flett, G. L., Hewitt, P. L., & Cheng, W. M. W. (2008). Perfectionism, distress, and irrational beliefs in high school students: Analyses with an abbreviated Survey of Personal Beliefs for adolescents. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, 26, 194-205.
- Flett, G. L., Hewitt, P. L., Besser, A., Su, C., Vaillancourt, T., Boucher, D., Munro, Y., Davidson, L. A., & Gale, O. (2016). The Child-Adolescent Perfectionism Scale: Development, psychometric properties, and associations with stress, distress, and psychiatric symptoms. *Journal of Psychoeducational Assessment*, 34, 634-652.
- Garner, D. M. (1991). *EDI-2. Eating disorder inventory-2. Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Hanstock, T. L., & O'Mahony, J. F. (2002). Perfectionism, acne, and appearance concerns. *Personality and Individual Differences*, 32, 1317-1325.
- Harris, D. L., & Carr, A. T. (2001). Prevalence of concern about physical appearance in the general population. *British Journal of Plastic Surgery*, 54, 223-226.
- Hartmann, A. S., Thomas, J. T., Greenberg, J. L., Matheny, N. L., & Wilhelm, S. (2014). A Comparison of Self-Esteem and Perfectionism in Anorexia Nervosa and Body Dysmorphic Disorder. *The Journal of Nervous and Mental Disease*, 202, 883-888.
- Hewitt, E. L., & Flett, G. L. (1991). Perfectionism in the self and social contexts: Conceptualization, assessment, and association with psychopathology. *Journal of Personality and Social Psychology*, 60, 456-470.
- Hewitt, P. L., Caelian, C. F., Chen, C., & Flett, G. L. (2014). Perfectionism, stress, daily hassles, hopelessness, and suicide potential in depressed psychiatric adolescents. *Journal of Psychopathology and Behavioral Assessment*, 36, 663-674.
- Hewitt, P. L., Caelian, C. F., Flett, G. L., Sherry, S. B., Collins, L., & Flynn, C. A. (2002). Perfectionism in children: Associations with depression, anxiety, and anger. *Personality and Individual Differences*, 32, 1049-1061.
- Hewitt, P. L., Newton, J., Flett, G. L., & Callander, L. (1997). Perfectionism and suicide ideation in adolescent psychiatric patients. *Journal of Abnormal Child Psychology*, 25, 95-101.

- Lambrou, C., Veale, D., & Wilson, G. (2012). Appearance concerns comparisons among persons with body dysmorphic disorder and nonclinical controls with and without aesthetic training. *Body Image*, 9, 86-92.
- Lapointe, L., & Emond, C. (2005). Relationships between suicidal behaviors and perfectionism dimensions in a non-clinical sample of adolescents. *Revue Quebecoise de Psychologie*, 26, 259-276.
- Leone, E. M. & Wade, T. D. (2018). Measuring perfectionism in children: a systematic review of the mental health literature. *European Child & Adolescent Psychiatry*, 27, 553-567.
- Libby, S., Reynolds, S., Derisley, J., & Clark, S. (2004). Cognitive appraisals in young people with obsessive-compulsive disorder. *Journal of Child Psychology and Psychiatry*, 45, 1076-84.
- Marques, L., LeBlanc, N., Weingarden, H., Greenberg, J. L., Traeger, L. N., Keshaviah, A. & Wilhelm, S. (2011). Body dysmorphic symptoms: Phenomenology and ethnicity. *Body Image*, 8, 163-167.
- Mastro, S., Zimmer-Gembeck, M. J., Webb, H. J., Farrell, L. & Waters, A. (2016). Young adolescents' appearance anxiety and body dysmorphic symptoms: social problems, self-perceptions and comorbidities. *Journal of Obsessive-Compulsive and Related Disorders*, 8, 50-55.
- Mayville, S. B., Katz, R., Gipson, M. T., & Cabral, K. (1999). Assessing the prevalence of body dysmorphic disorder in an ethnically diverse group of adolescents. *Journal of Child and Family Studies*, 8, 357-362.
- McCreary, B., Joiner, T., Schmidt, N., & Ialongo, N. (2004). The structure and correlates of perfectionism in African American Children. *Journal of Clinical Child & Adolescent Psychology*, 33, 313-324.
- McVey, G. L., Pepler, D., Davis, R., Flett, G. L., & Abdoell, M. (2002). Risk and protective factors associated with disordered eating during early adolescence. *Journal of Early Adolescents*, 22, 75-95.

- Mitchell, J. H., Newall, C., Broeren, S., & Hudson, J. L. (2013). The role of perfectionism in cognitive behaviour therapy outcomes for clinically anxious children. *Behaviour Research and Therapy*, 51, 547-554.
- Möllmann, A., Dietel, F. A., Hunger, A., & Buhlmann, U. (2017). Prevalence of body dysmorphic disorder and associated features in German adolescents: A self-report survey. *Psychiatry Research*, 254, 263–267.
- Muris, P., Meesters, C., & Schouten, E. (2002). A brief questionnaire of DSM-IV-defined anxiety and depression symptoms among children. *Clinical Psychology & Psychotherapy*, 9, 430–442.
- Neziroglu, F., & Barile, N. (2017). Environmental Factors in Body Dysmorphic Disorder. In K. A. Phillips (Ed.), *Body Dysmorphic Disorder: Advances in Research and Clinical Practice* (pp. 277-284). New York: Oxford University Press.
- O'Connor, R., Dixon, D., & Rasmussen, S. (2009). The structure and temporal stability of the child and adolescent perfectionism scale. *Psychological Assessment*, 21, 437–443.
- O'Connor, R. C., Rasmussen, S., & Hawton, K. (2010). Predicting depression, anxiety and self-harm in adolescents: The role of perfectionism and acute life stress. *Behaviour Research and Therapy*, 48, 52-59.
- Pamies, L., & Quiles, Y. (2014). Perfectionism and risk factors for the development of eating disorders in Spanish adolescents of both genders. *Anales de Psicologia*, 30, 620-626.
- Phillips, K. A. (2001). Body dysmorphic disorder. In: K. A., Phillips (Ed.), J. M., Oldham, & M. B., Riba, (Series Eds.), *Somatoform and Factitious Disorders: Review of Psychiatry* (pp. 67–94). Publishing Washington, DC: American Psychiatric Association.
- Phillips, K. A. (2004). Body dysmorphic disorder: recognizing and treating imagined ugliness. *World Psychiatry*, 3, 12-17.
- Phillips, K. A. (2005). *The broken mirror: Understanding and treating body dysmorphic disorder*. New York, NY: Oxford University Press.
- Phillips, K. A., Didie, E. R., Menard, W., Pagano, M. E., Fay, C., & Weisberg, R. B. (2006). Clinical features of body dysmorphic disorder in adolescents and adults. *Psychiatry Research*, 141, 305–314.

- Richardson, S. M. & Paxton, S. J. (2010). An evaluation of a body image intervention based on risk factors for body dissatisfaction: A controlled study with adolescent girls. *International Journal of Eating Disorders*, 43, 112–122.
- Rief, W., Buhlmann, U., Wilhelm, S., Borkenhagen, A., & Brähler, E. (2006). The prevalence of body dysmorphic disorder: a population-based survey. *Psychological Medicine*, 36, 877-885.
- Rosenblum, G. D., & Lewis, M. (1999). The relations among body image, physical attractiveness, and body mass in adolescence. *Child development*, 70, 50-64.
- Rudiger, J. A., Cash, T. F., Roehrig, M., & Thompson, K. (2007). Day-to-day body-image states: Prospective predictors of intra-individual level and variability. *Body Image*, 4, 1-9.
- Saling, M., Ricciardelli, L. A., & McCabe, M. P. (2005). A Prospective Study of Individual Factors in the Development of Weight and Muscle Concerns among Preadolescent Children. *Journal of Youth and Adolescence*, 34, 651–661.
- Schieber, K., Kollei, I., deZwaan, M., Müller, A., & Martin, A. (2013). Personality traits as vulnerability factors in body dysmorphic disorder. *Psychiatry Research*, 210, 242–246.
- Schneider, S. C., Baillie, A. J., Mond, J., Turner, C. M., & Hudson, J. L. (2016<sub>b</sub>). Measurement Invariance of a Body Dysmorphic Disorder Symptom Questionnaire across Sex: The Body Image Questionnaire–Child and Adolescent Version. *Assessment*, 1-10. doi:10.1177/1073191116679504.
- Schneider, S. C., Mond, J., Turner, C. M., & Hudson, J. L. (2017<sub>a</sub>). Subthreshold Body Dysmorphic Disorder in Adolescents: Prevalence and Impact. *Psychiatry Research*, 251, 125–130.
- Schneider, S. C., Mond, J., Turner, C. M., & Hudson, J. L. (2017<sub>b</sub>). Sex Differences in the Presentation of Body Dysmorphic Disorder in a Community Sample of Adolescents. *Journal of Clinical Child & Adolescent Psychology*, 1-13. doi: 10.1080/15374416.2017.1321001.
- Schneider, S. C., Turner, C. M., Mond, J., & Hudson, J. L. (2016<sub>a</sub>). Prevalence and correlates of body dysmorphic disorder in a community sample of adolescents. *Australian and New Zealand Journal of Psychiatry*, 51, 595-603.



- Shaw, H. E., Stice, E., & Springer, D. W. (2004). Perfectionism, Body Dissatisfaction, and Self-Esteem in Predicting Bulimic Symptomatology: Lack of Replication. *International Journal of Eating Disorders*, 36, 41-47.
- Sherry, S. B., Vriend, J. L., Hewitt, P. L., Sherry, D. L., Flett, G. L., & Wardrop, A. A. (2009). Perfectionism dimensions, appearance schemas, and body image disturbance in community members and university students. *Body Image*, 6, 83-89.
- Soreni, N., Streiner, D., McCabe, R., Bullard, C., Swinson, R., Greco, A., Pires, P., & Szatmari, P. (2014). Dimensions of perfectionism in children and adolescents with obsessive-compulsive disorder. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 23, 136-141.
- Steiger, H., Leung, F. Y. K., Puentes-Neuman, G., & Gottheil, N. (1992). Psychosocial profiles of adolescent girls with varying degrees of eating and mood disturbances. *International Journal of Eating Disorders*, 11, 121-131.
- Tylka, T. L. (2012). Positive Psychology Perspectives on Body Image. *Encyclopedia of Body Image and Human Appearance*, 2, 657-663.
- Veale, D. (2009). *Body Image Questionnaire–Child and Adolescent Version*. Retrieved from <http://www.kcl.ac.uk/ioppn/depts/psychology/research/ResearchGroupings/CADAT/Research/Body-Image-Questionnaires.aspx>.
- Veale, D. (2004). Advances in a cognitive behavioural model of body dysmorphic disorder. *Body Image*, 1, 113-125.
- Veale, D., & Neziroglu, F. (2010). *Body Dysmorphic Disorder: a treatment manual*. Wiley: Chichester.
- Veale, D., Boockock, A., Gournay, K., Dryden, W., Shah, F., Wilson, R., & Walbrun, J. (1996). Body Dysmorphic Disorder – A survey of fifty cases. *British Journal of Psychiatry*, 169, 196-201.
- Veale, D., Gledhill, L. J., Christodoulou, P., & Hodsoll, J. (2016). Body dysmorphic disorder in different settings: A systematic review and estimated weighted prevalence. *Body Image*, 18, 168–186.

- Veale, D., Kinderman, P., Riley, S., & Lambrou, C. (2003). Self-discrepancy in body dysmorphic disorder. *British Journal of Clinical Psychology, 42*, 157 – 169.
- Webb, H. J., Zimmer-Gembeck, M. J., Mastro, S., Farrell, L. J., Waters, A., & Lavell, C. H. (2015). Young adolescents' body dysmorphic symptoms: Associations with same- and cross-sex peer teasing via appearance-based rejection sensitivity. *Journal of abnormal child psychology, 43*, 1161-1173.
- Wilhelm, S., & Neziroglu, F. (2002). Cognitive theory of body dysmorphic disorder. In R. O. Frost, & G. Steketee (Eds.), *Cognitive Approaches to Obsessions and Compulsions*, (pp. 203–214). Amsterdam: Pergamon/Elsevier Sci.
- Wilhelm, S., Phillips, K. A., & Steketee, G. (2012). *Cognitive-behavioral therapy for body dysmorphic disorder*. New York: The Guilford Press.
- Ye, H. J., Rice, K. G., & Storch, E. A. (2008). Perfectionism and peer relations among children with obsessive-compulsive disorder. *Child Psychiatry and Human Development, 39*, 415-26.
- Zimmer-Gembeck, M. J., Webb, H. J., Farrell, L. J., & Waters, A. M. (2017). Girls' and boys' trajectories of appearance anxiety from age 10 to 15 years are associated with earlier maturation and appearance-related teasing. *Development and Psychopathology, 1-14*.

## 6. APPENDIX

### 6.1. Ethnic Composition of Sample from Study One

<b>Ethnicity Group</b>	<b>Year 7</b>	<b>Year 9</b>	<b>Year 11</b>	<b>Total</b>
Any Other Asian Background	0.4%	0.4%	1.3%	<b>0.7%</b>
Any Other Black Background	0.0%	0.0%	0.4%	<b>0.1%</b>
Any Other Ethnic Group	0.0%	0.0%	0.8%	<b>0.3%</b>
Any Other Mixed Background	1.7%	4.1%	1.7%	<b>2.5%</b>
Any Other White Background	5.0%	3.7%	2.9%	<b>3.9%</b>
Bangladeshi	0.0%	0.8%	0.8%	<b>0.6%</b>
Black - African	2.9%	0.8%	2.1%	<b>2.0%</b>
Black Caribbean	2.5%	0.8%	0.4%	<b>1.3%</b>
Chinese	0.0%	1.2%	0.0%	<b>0.4%</b>
Indian	0.8%	1.7%	0.4%	<b>1.0%</b>
Information not yet obtained	0.4%	0.0%	0.0%	<b>0.1%</b>
Pakistani	0.0%	0.0%	0.0%	<b>0.0%</b>
Refused	1.7%	0.8%	0.4%	<b>1.0%</b>
White - British	78.2%	81.1%	83.6%	<b>81.0%</b>
White - Irish	0.8%	0.4%	0.4%	<b>0.6%</b>
White and Asian	1.3%	1.7%	0.8%	<b>1.3%</b>
White and Black African	0.4%	0.4%	0.4%	<b>0.4%</b>
White and Black Caribbean	3.8%	2.1%	3.4%	<b>3.1%</b>
<b>Total</b>	<b>100.0%</b>	<b>100.0%</b>	<b>100.0%</b>	<b>100.0%</b>

## 6.2. Body Image Questionnaire – Child and Adolescent version (BIQ-C; Veale, 2009)

**We would like you to think about the 3 body features that you are the most unhappy with (or not completely satisfied with).** For example, this can include features like your nose, ears, teeth, feet, arms, skin, breasts, bum, genitals, facial hair, body hair, and the list can go on!



A body feature can be any part of your body, but remember, **do not** include any answers where the reason you are unhappy with them is because you think it is too fat or too skinny (i.e. related to body weight or fat).

**Put the feature that you are MOST unhappy with next to number 1 in the table below. Then the feature you are second most unhappy with next to number 2, and finally, your third most unhappy body part next to number 3.**

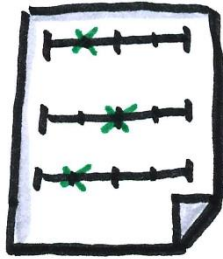
**In the column titled “What I don’t like about it”, please put your reason why you are unhappy/not satisfied with this body part.** Examples of some reasons are: because it is too big, too crooked, too small, too dark, too light, too long, too pointy, too flat, too wide, too lumpy, not smooth... etc.

	Body Feature	What I don’t like about it
1.		
2.		
3.		

**Please look at the questions below and answer them as honestly and accurately as possible. They refer to the features you identified above, unless otherwise stated.**

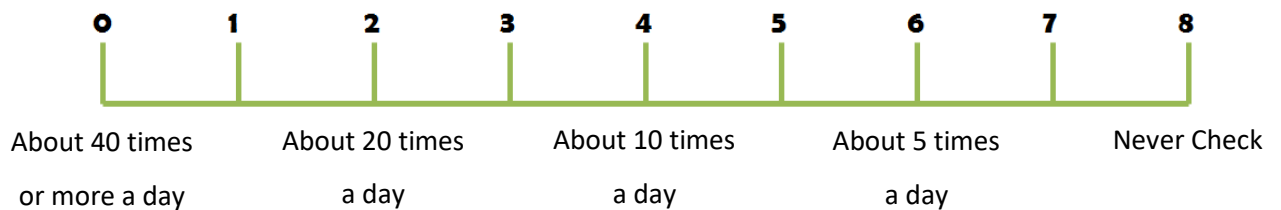
- 1) On an **average day**, how many minutes or hour(s) do you currently spend thinking about your feature(s)? Please add up all the time that your features are at the forefront of your mind and make the best estimate.

\_\_\_\_\_ minutes a day      **OR**      \_\_\_\_\_ hour(s) a day

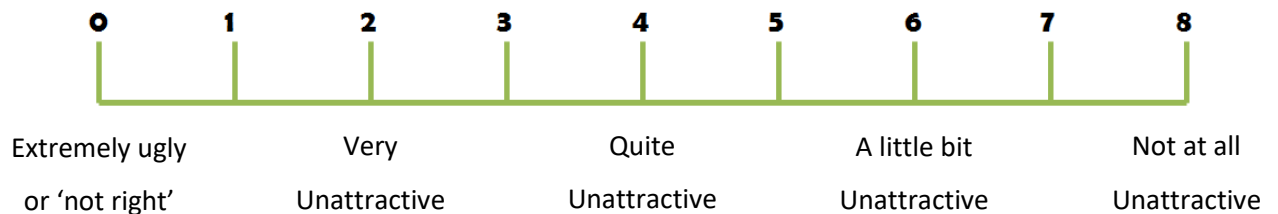


*Please read the next set of questions below carefully and circle/cross the number that best describes the way that you feel about your feature(s). Please read the labels carefully to ensure you are circling the number that describes how you feel.*

2) How often do you check your feature(s)? This is how often you check it on purpose, **not accidentally catch sight of it**. Please include looking at your feature in a mirror or other reflective surfaces (e.g. a shop window), taking photos, or looking at it directly or feeling it with your fingers.



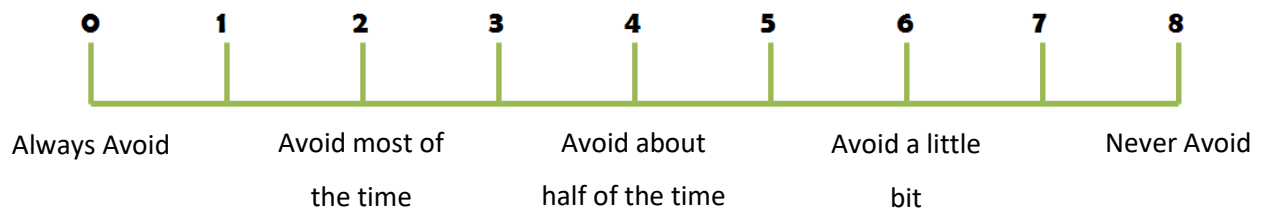
3) How much do you feel your feature(s) is ugly, unattractive or 'not right'?



4) How much does your feature(s) cause you a lot of distress (this is how much it upsets you when you think about it or see it)?



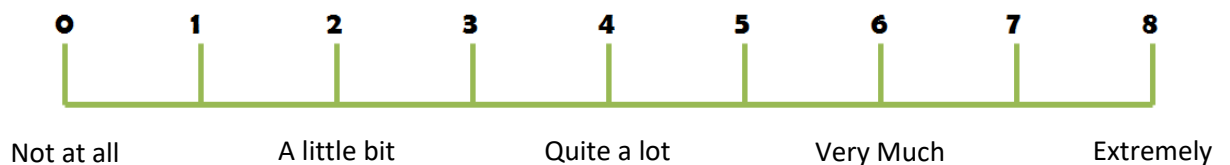
- 5) How often does your feature(s) lead you to avoid places or activities (e.g. things you might do with friends or family, places you may go or things that you might do at school)?



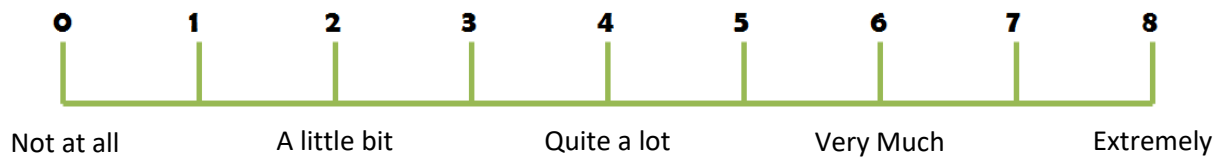
- 6) How much is your feature(s) on your mind? That is, you think about it a lot and it is hard to stop thinking about it?



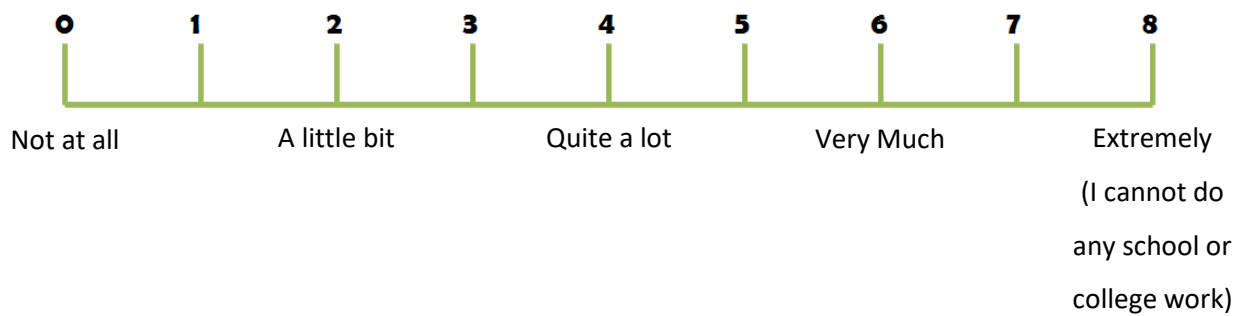
- 7) **If you have a girlfriend or boyfriend**, how much does your feature(s) have an effect on your relationship with him or her? Or **if you do not have a girlfriend or boyfriend but would like one**, how much does it have an effect on you getting one?



**8)** How much does your feature(s) have an effect on your relationship with your family (e.g. how much you enjoy doing things together or the number or arguments you have with them)?



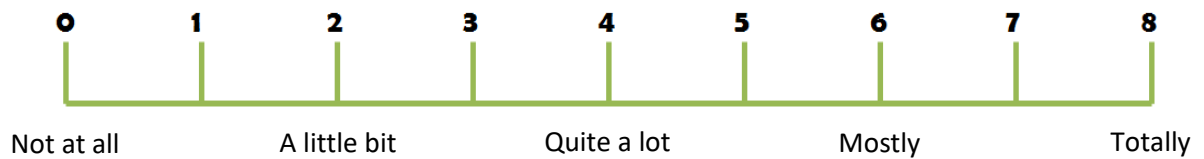
**9)** How much does your feature(s) get in the way with your school or college work?



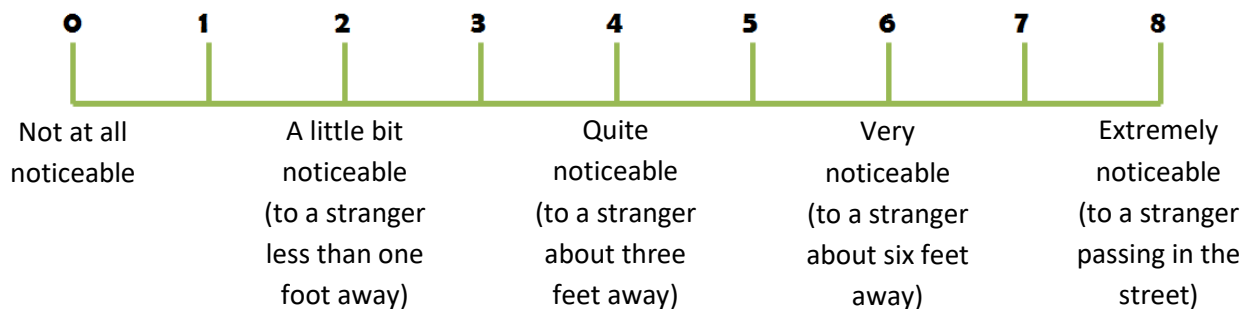
**10)** How much does your feature(s) get in the way with your social life (i.e. spending time with friends, going to parties)?



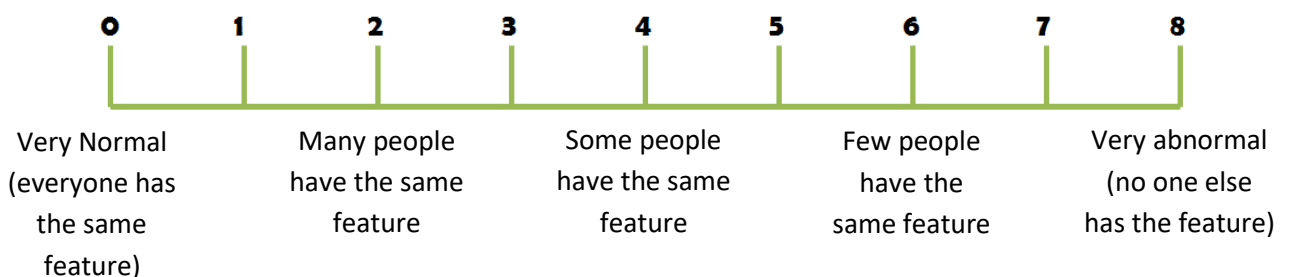
11) How much do you feel your appearance is the most important thing about you?



12) How noticeable do you feel your feature is to other people (if you do not cover it up e.g. using clothes, padding and/or makeup, and when the feature has not been pointed out to them)? **Please answer this question for the feature you ranked as #1 only (i.e. the feature you are most worried about).**



13) How does your feature compare to others of the same age, sex, and ethnic group? **Please answer this question for the feature you ranked as #1 only (i.e. the feature you are most worried about).**





### 6.3. Ethnic Composition of Sample from Study Two

<b>Ethnicity Group</b>	<b>Year 10</b>	<b>Year 11</b>	<b>Total</b>
Any Other Asian Background	0.4%	1.3%	<b>0.8%</b>
Any Other Black Background	0.4%	0.4%	<b>0.4%</b>
Any Other Ethnic Group	0.8%	0.8%	<b>0.8%</b>
Any Other Mixed Background	2.1%	1.7%	<b>1.9%</b>
Any Other White Background	4.2%	2.9%	<b>3.6%</b>
Bangladeshi	0.0%	0.8%	<b>0.4%</b>
Black - African	1.7%	2.1%	<b>1.9%</b>
Black Caribbean	0.8%	0.4%	<b>0.6%</b>
Chinese	0.0%	0.0%	<b>0.0%</b>
Indian	0.0%	0.4%	<b>0.2%</b>
Information not yet obtained	0.4%	0.0%	<b>0.2%</b>
Pakistani	0.0%	0.0%	<b>0.0%</b>
Refused	0.4%	0.4%	<b>0.4%</b>
White - British	83.3%	83.6%	<b>83.5%</b>
White - Irish	0.4%	0.4%	<b>0.4%</b>
White and Asian	2.5%	0.8%	<b>1.7%</b>
White and Black African	0.4%	0.4%	<b>0.4%</b>
White and Black Caribbean	2.1%	3.4%	<b>2.7%</b>
<b>Total</b>	<b>100.0%</b>	<b>100.0%</b>	<b>100.0%</b>

## 6.4. Child-Adolescent Perfectionism Scale (CAPS; Flett et al., 2016)

**Please read each sentence in the table below. We would like you to indicate how much you feel each statement is true about you. Please indicate your answer by marking (✓ or X) the one of the boxes numbered from “1” to “5.” The five possible answers for each sentence are listed below.**

- 1 = False - not at all true of me
- 2 = Mostly false
- 3 = Neither true nor false
- 4 = Mostly true
- 5 = Very true of me



*For example, if you were given the sentence, “I like to read comic books,” you would circle a “5” if this is very true of you. If you were given the sentence, “I like to keep my room neat and tidy”, you would circle a “1” if this was false and not at all true of you.*

	FALSE 1	2	Neither 3	4	TRUE 5
I try to be perfect in everything I do	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I want to be the best at everything I do	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
My parents don't always expect me to be perfect in everything I do	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I feel that I have to do my best all the time	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
There are people in my life who expect me to be perfect	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I always try for the top score on a test	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
It really bothers me when I don't do my best all the time	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
My family expects me to be perfect	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I don't always try to be the best	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
People expect more from me than I am able to give	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I get mad at myself when I make a mistake	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other people think I have failed if I do not do my very best all the time	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other people always expect me to be perfect	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I get upset if there is even one mistake in my work	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
People around me expect me to be great at everything	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
When I do something, it has to be perfect	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
My teachers expect my work to be perfect	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

I do not have to be the best at everything I do	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I am always expected to do better than others	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Even when I pass, I feel that I have failed if I didn't get one of the highest marks in the class	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I feel that people ask too much of me	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I can't stand to be less than perfect	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## 6.5. Revised Child Anxiety and Depression Scale – Short Version (RCADS-25; Ebesutani et al., 2012)

*Please put a circle around the word that shows how often each of these things happen to you. There are no right or wrong answers.*



<b>I feel sad or empty</b>	Never	Sometimes	Often	Always
<b>I worry when I think I have done poorly at something</b>	Never	Sometimes	Often	Always
<b>I would feel afraid of being on my own at home</b>	Never	Sometimes	Often	Always
<b>Nothing is much fun anymore</b>	Never	Sometimes	Often	Always
<b>I worry that something awful will happen to someone in my family</b>	Never	Sometimes	Often	Always
<b>I am afraid of being in crowded places (like shopping centres, the movies, buses, busy playgrounds)</b>	Never	Sometimes	Often	Always
<b>I worry what other people think of me</b>	Never	Sometimes	Often	Always
<b>I have trouble sleeping</b>	Never	Sometimes	Often	Always
<b>I feel scared if I have to sleep on my own</b>	Never	Sometimes	Often	Always
<b>I have problems with my appetite</b>	Never	Sometimes	Often	Always
<b>I suddenly become dizzy or faint when there is no reason for this</b>	Never	Sometimes	Often	Always
<b>I have to do some things over and over again (like washing my hands, cleaning or putting things in a certain order)</b>	Never	Sometimes	Often	Always
<b>I have no energy for things</b>	Never	Sometimes	Often	Always
<b>I suddenly start to tremble or shake when there is no reason for this</b>	Never	Sometimes	Often	Always
<b>I cannot think clearly</b>	Never	Sometimes	Often	Always
<b>I feel worthless</b>	Never	Sometimes	Often	Always
<b>I have to think of special thoughts (like numbers or words) to stop bad things from happening</b>	Never	Sometimes	Often	Always
<b>I think about death</b>	Never	Sometimes	Often	Always
<b>I feel like I don't want to move</b>	Never	Sometimes	Often	Always
<b>I worry that I will suddenly get a scared feeling when there is nothing to be afraid of</b>	Never	Sometimes	Often	Always
<b>I am tired a lot</b>	Never	Sometimes	Often	Always

<b>I feel afraid that I will make a fool of myself in front of people</b>	Never	Sometimes	Often	Always
<b>I have to do some things in just the right way to stop bad things from happening</b>	Never	Sometimes	Often	Always
<b>I feel restless</b>	Never	Sometimes	Often	Always
<b>I worry that something bad will happen to me</b>	Never	Sometimes	Often	Always

## 6.6. Ethical Approval Letter



Rachel Quinn

1 February 2017

Dear Rachel ,

Study Title: Understanding "normal" appearance concerns in British school children and how they are related to perfectionism.

Study Reference: Review Reference

I am pleased to inform you that full approval for your project has been granted by the PNM Research Ethics Subcommittee .

For your information, ethical approval has been granted for 3 years from 1 February 2017. If you need approval beyond this point, you will need to apply for an extension at least two weeks before this. You will be required to explain the reasons for the extension. However, you will not need to submit a full re-application unless the protocol has changed.

Ethical approval is required to cover the data-collection phase of the study. This will be until the date specified in this letter. However, you do not need ethical approval to cover subsequent data analysis or publication of the results. For secondary data-analysis, ethical approval is applicable to the data that is sensitive or identifies participants.

Please ensure that you follow the guidelines for good research practice as laid out in UKRIO's Code of Practice for research:  
<http://www.kcl.ac.uk/innovation/research/support/conduct/cop/index.aspx>.

Please note you are required to adhere to all research data/records management and storage procedures agreed to as part of your application. This will be expected even after the completion of the study.

If you do not start the project within three months of this letter, please contact the Research Ethics Office.

Please note that you will be required to obtain approval to modify the study. This also encompasses extensions to periods of approval. Please refer to the URL below for further guidance about the process:

<http://www.kcl.ac.uk/innovation/research/support/ethics/applications/modifications.aspx>

Please would you also note that we may, for the purposes of audit, contact you from time to time to ascertain the status of your research.

If you have any query about any aspect of this ethical approval, please contact the Research Ethics Office:

(<http://www.kcl.ac.uk/innovation/research/support/ethics/contact.aspx>)

We wish you every success with this work.

Yours sincerely,

James Patterson - Senior Research Ethics Officer

**For and on behalf of**

Dr Jane Petty, Chair of the PNM Research Ethics Subcommittee

Cc: Amita Jassi

## 6.7. Supplementary Statistical Analysis

### Effect of Gender on BIQ-C scores

#### **Independent T-Test:**

An independent-samples t-test was run to determine if there were differences in BIQ-C scores between males and females. There were 15 outliers in the data, as assessed by inspection of a boxplot, and these were removed from the analysis ( $N = 517$ , 266 male and 251 female participants). The assumption of homogeneity of variances was violated, as assessed by Levene's test for equality of variances ( $p < .001$ ). Female BIQ-C scores were a statistically significantly higher ( $M = 32.42$ ,  $SD = 15.98$ ) than male BIQ-C scores ( $M = 19.04$ ,  $SD = 11.81$ ) ( $t(459.089) = -10.777$ ,  $p < .001$ ).

#### **Mann-Whitney U Test (Non-parametric):**

A Mann-Whitney U test was run to determine if there were differences in BIQ-C score between males and females ( $N = 532$ ). Distributions of the BIQ-C scores for males and females were not similar, as assessed by visual inspection. BIQ-C scores for females (mean rank = 327.56) were statistically significantly higher than for males (mean rank = 211.54) ( $U = 50,668.000$ ,  $z = 8.694$ ,  $p < .001$ ).

### Effect of Age on BIQ-C scores

#### **One-Way ANOVA:**

A one-way ANOVA was conducted to determine if BIQ-C scores were different for groups with different ages. Participants were classified into three groups: Year 7 ( $N = 218$ ), Year 9 ( $N = 198$ ), and Year 11 ( $N = 114$ ). There were 4 outliers, as assessed by boxplot, which were removed, and the assumption of homogeneity of variances was violated, as assessed by Levene's test of equality of variances ( $p = .016$ ). The differences in BIQ-C score between these age groups was not statistically significant, Welch's  $F(2, 294.367) = 0.382$ ,  $p = .683$ .

#### **Kruskal-Wallis H Test (Non-parametric):**

A Kruskal-Wallis H test was run to determine if there were differences in BIQ-C score between the three age groups of participants, those currently studying in Year 7 ( $N = 222$ ), Year 9 ( $N = 198$ ), and Year 11 ( $N = 114$ ) at school. Distributions of BIQ-C scores were similar for all groups, as assessed by visual inspection of a boxplot. Median BIQ-C scores (median = 24.0, 21.0, and 25.5 respectively) were not statistically significantly different between groups ( $\chi^2(2) = 1.860$ ,  $p = .395$ ).

# VOLUME II

## *Clinical Case-Studies & Service-Related Project*

**Rachel Quinn**

May 2018

**Institute of Psychiatry, Psychology, and Neuroscience,  
King's College London**

Thesis submitted in partial fulfilment of the degree of Doctorate in Clinical Psychology



## ACKNOWLEDGEMENTS

---

Special thanks to all the people who agreed for me to write up their work as a case study. It was a pleasure to be able to work with such resilient clients and I am grateful for them sharing their stories with me. Despite facing many stressors, they were brave and tenacious, and it was an honour to be a part of their journey.

I am also especially grateful for the support from my supervisors in completing this series of work. Thank you, Dr Nick Dobson, Dr Suraba Mahendiran, Dr Becca Piper, Dr David Matthews, Dr Anna Redfern, and Dr Jane Ellis. I learnt something new from each of you, so thank you for sharing your wisdom and knowledge.

## TABLE OF CONTENTS

---

### CASE STUDY ONE

**Climbing the Fear Ladder:** CBT for needle phobia in an 18-year-old woman

*Page 4*

### CASE STUDY TWO

**Overcoming the Anger:** CBT with an 11-year-old boy

*Page 33*

### CASE STUDY THREE

**Detecting the Dysfunction:** A neuropsychological assessment of memory with a 74-year-old man diagnosed with Parkinson's disease

*Page 60*

### CASE STUDY FOUR

**Widening the Access to "CUES-Ed":** Using an adapted CBT school-based intervention for children with Social, Emotional, and Mental Health (SEMH) needs



*Page 84*

### SERVICE-RELATED PROJECT

**Promoting Recovery within Croydon IAPT:** Which clients do not recover?

*Page 113*

## CASE STUDY ONE

---

### **Climbing the Fear Ladder:**

CBT for needle phobia in an 18-year-old woman

---

**Supervised by:** Dr Nick Dobson

## ABSTRACT

---

This case discussion outlines the assessment, formulation, intervention, and evaluation of an 18-year-old woman with a specific phobia of needles. 'Sophie' was referred to the IAPT service by her GP for the assessment and treatment of a specific phobia of needles. At this time, Sophie was required to have an increasing number of injections which she could not avoid, which is why she chose to seek help. She received 12 sessions of Cognitive Behavioural Therapy (CBT) which consisted of goal-setting, psycho-education, formulation, development of a fear ladder, in-vivo exposure, and relapse prevention. Weekly outcome measures were used to monitor Sophie's mood and anxiety throughout therapy, which showed symptom improvement by the end of therapy. Sophie was also able to reach the top of her fear ladder by successfully having a flu vaccination. Limitations and reflections are also discussed.

# CONTENTS

---

<b>1. LITERATURE REVIEW.....</b>	<b>7</b>
1.1. Needle Phobia .....	7
1.2. Psychological Models and Intervention .....	7
1.2.1. Behavioural Model .....	7
1.2.2. Cognitive Model .....	8
1.2.3. Interventions .....	8
<b>2. CASE DESCRIPTION .....</b>	<b>10</b>
2.1. Background.....	10
2.2. Referral.....	10
2.3. Assessment.....	10
2.4. Outcome Measures .....	11
<b>3. FORMULATION.....</b>	<b>12</b>
<b>4. INTERVENTION .....</b>	<b>14</b>
4.1. Engagement .....	14
4.2. Session Structure.....	14
<b>5. OUTCOMES .....</b>	<b>18</b>
5.1. Behavioural Results .....	18
5.2. Outcome Measures .....	18
5.3. Goals.....	20
<b>6. DISCUSSION .....</b>	<b>21</b>
6.1. Conclusions and Limitations.....	21
6.2. Reflections.....	21
<b>7. REFERENCES .....</b>	<b>23</b>
<b>8. APPENDICES .....</b>	<b>26</b>
8.1. Goal-Setting Worksheet (used in Session 1) .....	26
8.2. Vicious Flower Formulation (used in Session 2).....	27
8.3. Fear Ladder Handout (used in Session 3).....	28
8.4. Virtuous Flower Formulation (used in Session 9) .....	30
8.5. Relapse Prevention Plan (used in Session 12).....	31
8.6. Reviewed Goals (used in Session 12) .....	32

# 1. LITERATURE REVIEW

---

## 1.1. Needle Phobia

Specific phobias have been estimated to have a lifetime prevalence of 7.7% in Europe and 12.5% in the U.S., and they are considered one of the most common anxiety disorders (Alonso et al., 2004; Kessler et al., 2005). The specific phobia of needles falls within the ‘blood-injection-injury’ classification of phobias (Curtis, Magee, Eaton, Wittchen, & Kessler, 1998), and is characterised by anxious and avoidant behaviour when presented with needles or associated stimuli. This phobia is described as somewhat different from others, primarily due to the presence of the ‘vasovagal reflex’ (a sudden slowing of the heart and drop in blood pressure), which in some instances leads to fainting (Willemsen, Chowdhury, & Briscall, 2002).

Needle phobia can easily go undetected, meaning that it may only be brought to a clinician’s attention after a client is in need of regular injections or an operation (Willemsen et al., 2002). Andrews and Shaw (2010) suggest that between 4-25% of the general population suffer from some degree of needle phobia, where you can assume that a fair proportion goes untreated. This can be potentially damaging to the individual’s physical health, due to their avoidance to seek help for health concerns.

## 1.2. Psychological Models and Intervention

### 1.2.1. Behavioural Model

The two-stage theory of fear development (Mowrer, 1939) suggests that phobias develop as a result of classical conditioning, whereby a neutral stimulus becomes associated with an aversive stimulus, which is then maintained through negative reinforcement. However, this approach has been described as insufficient due to it being unable to explain why many individuals fail to recall a specific conditioning event that led to the phobia, and many people who experience traumatic events do not necessarily go on to develop a phobia (Hofmann & Reinecke, 2010). Rachman (1978) added to this model by hypothesising that “social learning” could explain the acquirement of phobias through two pathways, and which do not require direct contact with the feared stimulus. That is (1) informational transmission (learning that an object/situation is dangerous from information transmitted by others e.g. television or parent), and (2) vicarious acquisition (observing another’s overtly fearful reaction to an object/situation). The addition of these pathways increased the robustness of this model; however, it still ignored the concept of

biological (evolutionary) ‘preparedness’ – the perspective that some individuals may be more biologically susceptible to the adverse experiences that lead to the development of anxiety disorders (Coelho & Purkis, 2009).

#### 1.2.2. Cognitive Model

The cognitive perspective of specific phobia expands on the behaviourist ideas, suggesting that it is the individual’s thinking that is responsible for the maintenance of the phobia, specifically the attributions given to a stimulus regarding its level of safety/danger, and the level of control the person has over the situation (Coelho & Purkis, 2009). The individual will therefore consider their anxiety response as rational to the situation (because of perception biases/interpretation/memory), and consequently their safety behaviours are seen as logical (Kirk & Rouf, 2004). There is no specific cognitive model available for specific phobia, although after considering the relevant theories, Kirk and Rouf (2004) suggest that the following processes should be recognised when understanding the development and maintenance of phobias:

- The development of fear and assumptions of increased vulnerability.
- The anxious thoughts/images about the trigger and anxiety symptoms (fear of fear).
- The overestimation of threat and negative consequences.
- The underestimation of coping and rescue factors.
- The maintenance of fear through anxious predictions/physiological arousal/hypervigilance/safety behaviours.
- Secondary cognitions in the form of negative self-judgements which can lead to depression, hopelessness, loss of confidence and low self-esteem (e.g. “I am weak”).

#### 1.2.3. Interventions

Specific phobia can be considered one of the most treatable disorders, with a vast number of interventions available (Wolitzky-Taylor, Horowitz, Powers & Telch, 2008). However, it is exposure therapy that is considered the most widely accepted treatment, which consists of exposing the individual to their feared object to allow them to learn that their negative predictions do not actually occur, thus weakening the fear response (Cisler, Lohr, Sawchuk, & Olatunji, 2010). The exposure should be the foundation of treatment, with additional elements added as appropriate (e.g. applied tension) (Hood & Antony, 2012). Exposures can be achieved through a number of variations, such as: in vivo, imagined, interoceptive, virtual reality, Eye

Movement Desensitization and Reprocessing (EMDR), applied tension, and applied relaxation (for details please refer to Hood & Antony, 2012).

Drawing on the psychological models of specific phobia outlined above, a cognitive behavioural framework can be helpful to target both the behavioural and cognitive components of the phobia. As cognitive factors are arguably central to the maintenance of the phobia, some researchers believe that exposure is only effective if cognitive change also occurs (Foa & Kozak 1986; Salkovskis 1991). Kirk and Rouf (2004) explain that through the identification of maladaptive cognitions, treatment can be precisely targeted and more efficient for the individual. This can then facilitate new understandings that the feared stimuli are not (or are unlikely to be) dangerous, and so escape, avoidance, and other safety-seeking behaviours are unnecessary. Techniques used to achieve this include cognitive restructuring and guided threat appraisal (Wolitzky-Taylor et al., 2008).

Regarding the evidence-base, three independent meta-analyses have been carried out investigating the effectiveness of psychological treatments for specific phobia (Barlow, Moscovitch & Micco, 2004; Choy, Fyer & Lipsitz, 2007; Wolitzky-Taylor et al., 2008) - all of which suggest that exposure-based treatment is the most effective. Additionally, Choy and colleagues (2007) found that specific psychological treatments have varying levels of efficacy for each of the phobia sub-types. For instance, blood-injection-injury phobia was found to respond well to applied muscle tension. Wolitzky-Taylor and colleagues (2008) however did not support this finding but did identify that multiple exposure sessions are more effective than one session, particularly at follow-up. Therefore, they recommend that clinicians should deliver treatment in multiple sessions to enhance long-term treatment gains.

Taken together, the key component in the treatment for specific phobia is exposure. CBT appears to be a holistic approach which allows a clinician to consider both the behavioural and cognitive aspects of the development and maintenance of the phobia, although research suggests that only the exposure aspect is absolutely necessary (Wolitzky-Taylor et al., 2008). However, treatment can be customised to the individual by incorporating additional elements where appropriate (Hood & Antony, 2012).



## 2. CASE DESCRIPTION

---

### 2.1. Background

'Sophie' (pseudonym) was an 18-year-old, White British female, who lived with her parents and younger brother in a borough of London. She was a full-time Student studying for her A-Levels.

### 2.2. Referral

Sophie was referred to an 'Improving Access to Psychological Therapies' (IAPT) service by her GP for the assessment and treatment of a specific phobia of needles. At this time, Sophie was required to have an increasing number of injections which she could not avoid, which is why she chose to seek help.

### 2.3. Assessment

Following her referral, Sophie received a telephone triage assessment and, as deemed suitable for the service, she was allocated to the appropriate stepped-care model for treatment. In this case it was the step three waiting list for CBT, as well as the 'Getting Started Workshop' – a one-off informal group to give clients information about the service, how to prepare, and what they can do whilst waiting for therapy. After being on the waiting list for approximately five months, Sophie attended a one-hour, face-to-face assessment session prior to starting treatment. The aim was to obtain a first-hand account of Sophie's current problems to date and to understand how this was interfering with her life.

Sophie reported that since the age of 14, she had become extremely anxious when in the presence of needles, including talking about needles and seeing them on television. At these times, Sophie described physiological responses, that is: feeling faint (but had never fainted), an increased heart rate, sweaty hands, and tension. She also reported experiencing negative images of herself being injected and thoughts such as: "What if the needle snaps" and "I should be over this by now". Sophie described feelings of anxiety and fear, but also anger and embarrassment for responding this way. Sophie felt that her problem did not impact on many areas of her life, as she did not frequently come across needles, and so generally coped well day-to-day at home/school, and frequently engaged in activities that she enjoyed.

Confidentiality and the boundaries of therapy were explained, as well as assessing her current level of risk. No risk issues were identified as Sophie expressed no thoughts, intentions, or plans

to harm herself or others. The expectations and structure of therapy were discussed, and the basic theoretical components of CBT.

## 2.4. Outcome Measures

In addition to Sophie's qualitative account, quantitative outcome measures were used to assess the extent of the problem. These were:

- Patient Health Questionnaire - 9 (PHQ-9; Kroenke, Spitzer, & Williams, 2001) - symptoms of low mood/depression;
- Generalized Anxiety Disorder - 7 (GAD-7; Spitzer, Kroenke, Williams, & Löwe, 2006) - symptoms of general anxiety;
- IAPT Phobia Scale – avoidance of a specific object/situation;
- Work and Social Adjustment Scale (W & SAS; Mundt, Marks, Shear, & Greist, 2002) - problems in functioning with work/home management/social leisure activities/private leisure activities/family and relationships.

These measures have shown to have good reliability and validity for measuring symptom severity (Kroenke et al., 2001; Spitzer et al., 2006). As per the service's guidelines, these measures are used weekly to monitor mood and anxiety symptoms and assist in determining patient recovery (by classifying an individual's symptoms from healthy to severe).

At this assessment, Sophie scored two on the PHQ-9 indicating no clinical symptoms of depression; 10 on the GAD-7 indicating 'Moderate' symptoms of anxiety; five out of eight on the specific phobia scale, suggesting that Sophie did actively avoid needles; and zero on the W & SAS suggesting that this problem did not impact on her relationships, work, home, or social life.

### 3. FORMULATION

---

Following the assessment, a formulation was hypothesised to understand Sophie's presenting problem, and this was continually added to throughout therapy. A cross-sectional model by Greenberger and Padesky (1995) was used to map out the presenting problem in the context of the 'Five Ps' model (Dudley & Kuyken, 2006).

Sophie presented with a phobia of needles/injections, which resulted in a significant physiological response and avoidant behaviour. She would interpret the situation as dangerous by believing that the needle would cause her significant harm, and would consequently feel fear, as well as anger and embarrassment for her behaviour.

In terms of the predisposing factors, Sophie reported no previous mental health difficulties within her family, however she did describe her family as having "anxious personalities". This is likely to have made Sophie more vulnerable to developing an anxious character, which admittedly she believed she had. This could also be supported by her report of being treated for separation anxiety as a child. Moreover, Sophie's memory of her mother and younger brother being distressed when receiving an injection may have triggered the onset of the phobia. It appears as though certain behaviours were maintaining the anxiety. For instance, Sophie would avoid/escape from injections; become hyper-vigilant for needles; worry before medical appointments; and seek reassurance from her family. These subsequently had a negative impact on her thoughts and fed back into the anxiety.

Positively, Sophie did appear to have the motivation to change and her family were supportive of her seeking help. Additionally, Sophie was able to have a successful series of dental injections recently. Please see Figure 1 for a diagram of Sophie's formulation.

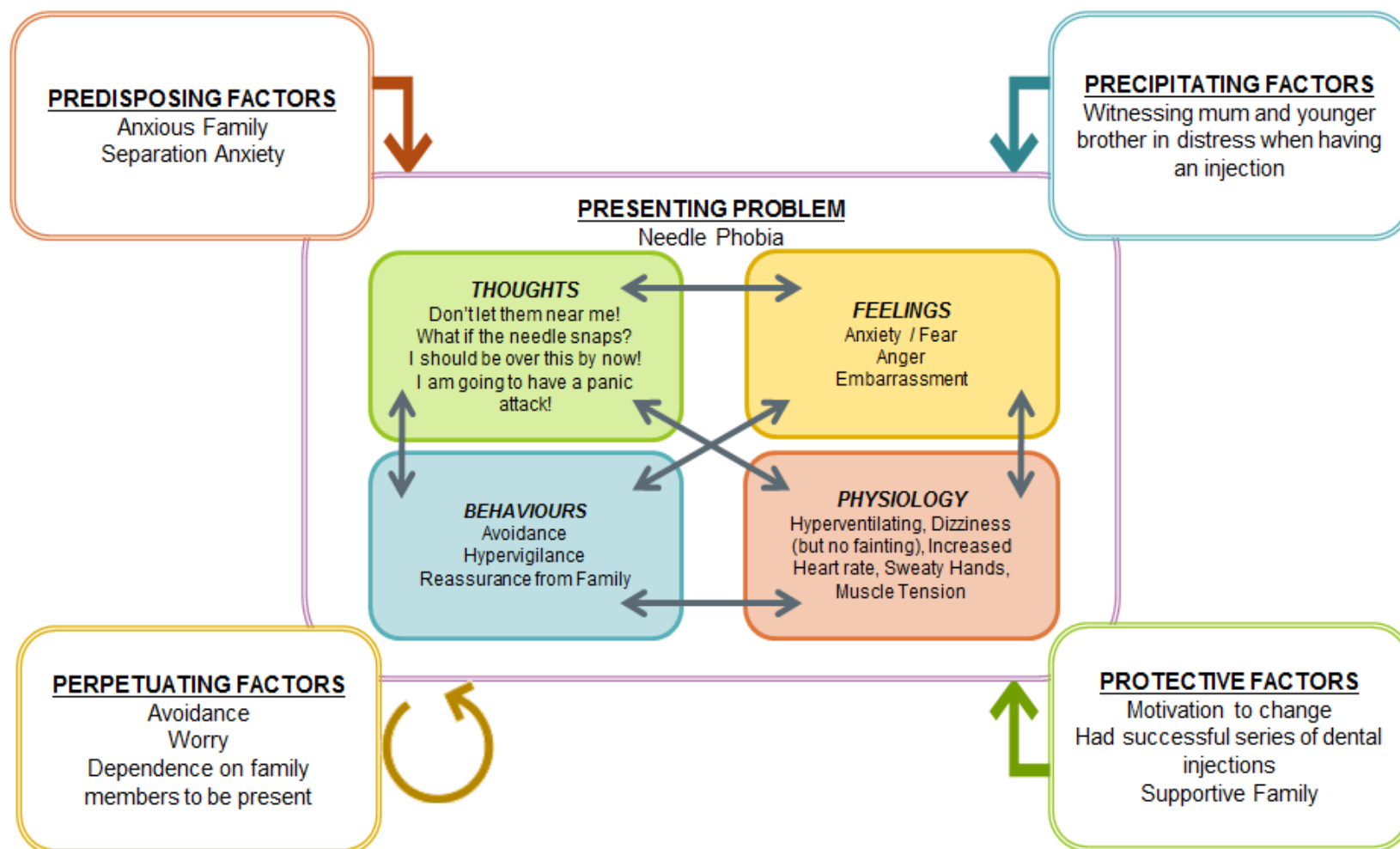


Figure 1: A cross-sectional formulation of Sophie's presenting problem, in the context of the five P's model.

## 4. INTERVENTION

---

Due to the evidence-base demonstrating CBT to be an effective treatment for the presenting problem, and Sophie appearing motivated and understanding this model, CBT was the provided treatment for this case. The 'applied tension' technique developed by Öst and Sterner (1986) was also considered and deemed unnecessary due to the absence of fainting.

Treatment comprised of 12 one-hour weekly sessions. Each session followed a CBT framework of: setting the session agenda, bridging from the previous session, reviewing any homework, carrying out the collaboratively set session activities, and setting new homework. There was also an opportunity for Sophie to provide any feedback at the end of each session, as well as to confirm the next appointment.

### 4.1. Engagement

Throughout therapy Sophie engaged well. She actively participated in all the activities/tasks set, both in and outside of sessions, despite these being highly anxiety-provoking. Moreover, Sophie appeared to utilise the sessions where necessary, for example she was honest when she had other concerns on her mind, and appropriately put these on our agenda to discuss (e.g. her relationship worries). It was always clear that Sophie was actively listening, as when asked to feedback key messages she had learnt, these were reflective of the session content and she would make links between this and her learning from other sessions/school. There were times when Sophie would arrive up to 10 minutes late to the session; however, she always took responsibility for this and would inform me as early as possible via email if this were to happen. It did not appear that this was avoidance relating to her phobia, but rather that Sophie relied on family members to drive her to the session and admitted finding it difficult to wake up for the early appointment.

### 4.2. Session Structure

#### **Session 1: Goal-Setting and Psycho-Education**

The first treatment session was used to collaboratively make a plan for treatment and set 'SMART' therapy goals (Specific, Measurable, Achievable, Realistic and Time limited) (Please see Appendix 8.1 for Sophie's Goal-Setting Worksheet). Time was also spent normalising Sophie's

phobia by discussing the helpful functions of anxiety and the 'Fight or Flight' response, as well as considering 'Unhelpful Thinking Styles'.

During the assessment, Sophie had expressed some concerns with her own general worrying, and so time was also spent exploring/clarifying this and whether it would be a treatment target. However, Sophie chose to primarily focus on her phobia, and we agreed that if any worries occurred we could put this on our agenda to discuss.

### **Session 2: Formulation of a Specific Example**

A 'vicious flower' model, based on the most recent situation of when Sophie was exposed to a needle, was used to elicit Sophie's thoughts, feelings, physiology, and behaviours, and identify the links between them. This model was chosen rather than the formulation discussed earlier, as it was felt that this model was more client-friendly in demonstrating the role of maintaining factors and in providing a rationale for therapy. Whereas the previous model was more comprehensive which allowed the therapist to be clear on the presenting problem, to consider the suitability for therapy, and to reflect on any considerations required for therapy.

Sophie found it difficult to think about what thoughts she was having in this situation to begin with, and so we discussed what imagery she experienced to help access these – which Sophie found helpful. A downward arrow questioning style was also used to help understand what types of negative scenarios or threat appraisals these thoughts represented. A copy of Sophie's formulation can be found in Appendix 8.2.

### **Session 3: Exposure Rationale and Fear Ladder**

The rationale for undergoing exposure tasks was discussed using guided discovery, allowing Sophie to consider what she would ask a friend to do if she was to hypothetically help them overcome a phobia of dogs. Sophie was able to independently explain the basic elements of exposure, which were then applied to her own phobia. Anxiety graphs were also utilised to demonstrate the changes in anxiety when exposed to a feared object compared to when it is avoided. Sophie then developed a 'Fear Ladder' which ranked all the different situations that trigger the anxiety, from least anxiety-provoking to most. This allowed her to develop a systematic approach for the exposures, guiding the treatment in terms of where the exposures should begin, and what she would hope to achieve by the end. A copy of Sophie's Fear Ladder can be found in Appendix 8.3.

### **Sessions 4-11: Exposure Tasks**

In the succeeding eight sessions, exposure tasks were completed covering a range of scenarios on Sophie's Fear Ladder, eventually concluding with Sophie having a flu vaccination in Session 11. These exposures were not necessarily completed in order (although to ensure engagement we began with exposures at the bottom), namely as the literature recommends that to maximise exposure, tasks should be variable and violate the client's expectancies (Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014).

Similarly, not every scenario could be completed within sessions due to the resources available, and so Sophie also continued with these as homework. Sophie often completed spontaneous exposure tasks between sessions, which was a credit to her dedication, but sometimes made it difficult to prepare appointments as she would often have completed the exposure planned for the subsequent session. Therefore, I always ensured I had two options for the exposures (either being two different tasks on her ladder, or two levels of the same task e.g. pictures of isolated needles vs. pictures of needles injecting the skin).

The conditions of the exposure were also stipulated prior to each task, as well as afterwards, to ensure Sophie understood how to get the most out of her treatment (i.e. that no safety behaviours should be used during the exposure and each exposure should be repeated as much as possible) – Sophie was soon able to recite these herself when prompted. As part of the exposures, Sophie was required to predict what the consequences of the exposure would be, which were then monitored during the exposure, allowing Sophie to reflect on the comparisons between the predicted and actual response. Moreover, Sophie was encouraged to identify the thoughts and feelings she was experiencing; to consider what she had learnt from each exposure experience; and reflect how this was developing with each exposure task. Table 1 outlines the actual exposure tasks completed throughout therapy and distinguishes whether these were within the session or set as homework.

Alongside the exposure work, time was also spent engaging in other therapeutic tasks. For instance, at session 6 Sophie's goals were reviewed; during session 9 a 'Virtuous Flower' model was used to develop a positive formulation (see Appendix 8.4); and we discussed the use of imagery and mindfulness as a coping strategy.

**Table 1: A list of each exposure task completed as part of therapy and in what setting this occurred.**

Exposure Task	Setting
Pictures of isolated needles.	Session 4 and set as Homework to repeat
Video of a medical programme.	Spontaneous Homework
Video of a needle injecting the heart in a medical programme.	Session 5
To have a conversation with someone about needles.	Homework session 5
Actual needle on a table (which Sophie subsequently interacted with out of choice).	Session 6
To take the needle home and continue interacting with it (risk/safety discussed).	Homework session 6
Touching the tip of the needle and putting it close to her face.	Spontaneous Homework
Touching a needle against various body parts whilst looking at pictures of needles injecting skin.	Session 7 and set as Homework to repeat
Video of a bone marrow transplant.	Spontaneous Homework
Discussing injections with an IAPT colleague (trained nurse) / watching colleague inject herself (for personal medical reasons) / watching colleague bend the needle to show it does not snap.	Session 8
Imaginary exposure of colleague injecting self / discussing injecting self with diabetic friend / pushing the edge of a needle (risk/safety discussed).	Homework session 8
To book an appointment for a flu vaccination.	Session 9 homework
Flu Vaccination.	Session 11

## Session 12: Relapse Prevention

In the final session a 'Relapse Prevention Plan' was collaboratively created to include what Sophie had learnt from therapy, the maintaining factors, her coping strategies, what she finds unhelpful, what will keep her well in future, and what she can do if she experiences a setback. A copy of this can be found in Appendix 8.5. Time was also spent reviewing Sophie's goals (Appendix 8.6), and her weekly outcome scores using visual graphs (from triage through to discharge). Sophie reported finding this task very beneficial in seeing her progress visually over the previous months. These can be found in the following '5.2 Outcomes' section (Figure 2).



## 5. OUTCOMES

### 5.1. Behavioural Results

At the start of therapy Sophie reported not being able to speak about needles/injections without experiencing significant anxiety. It was physically evident in the exposure tasks that Sophie was experiencing anxiety, yet to her credit she pursued with the task and this response gradually diminished. Similarly, Sophie became more confident with the exposure tasks, eventually booking her own appointment to have a flu vaccination. This was primarily due to her high level of motivation, but also the collaborative development of a formulation to understand how the phobia was being maintained; explaining the role of anxiety; and providing a sound rationale for exposure. Sophie acknowledged that she never thought she would be able to “do something like this” - that is, actively engage with something she had feared for so long. During the vaccination Sophie was slightly hesitant (pulling her arm away from the pharmacist) but was able to draw on techniques learnt to help cope with the situation and successfully complete the exposure. Sophie reported that this was the “quickest” she had ever had an injection (from hours to approximately 10 minutes).

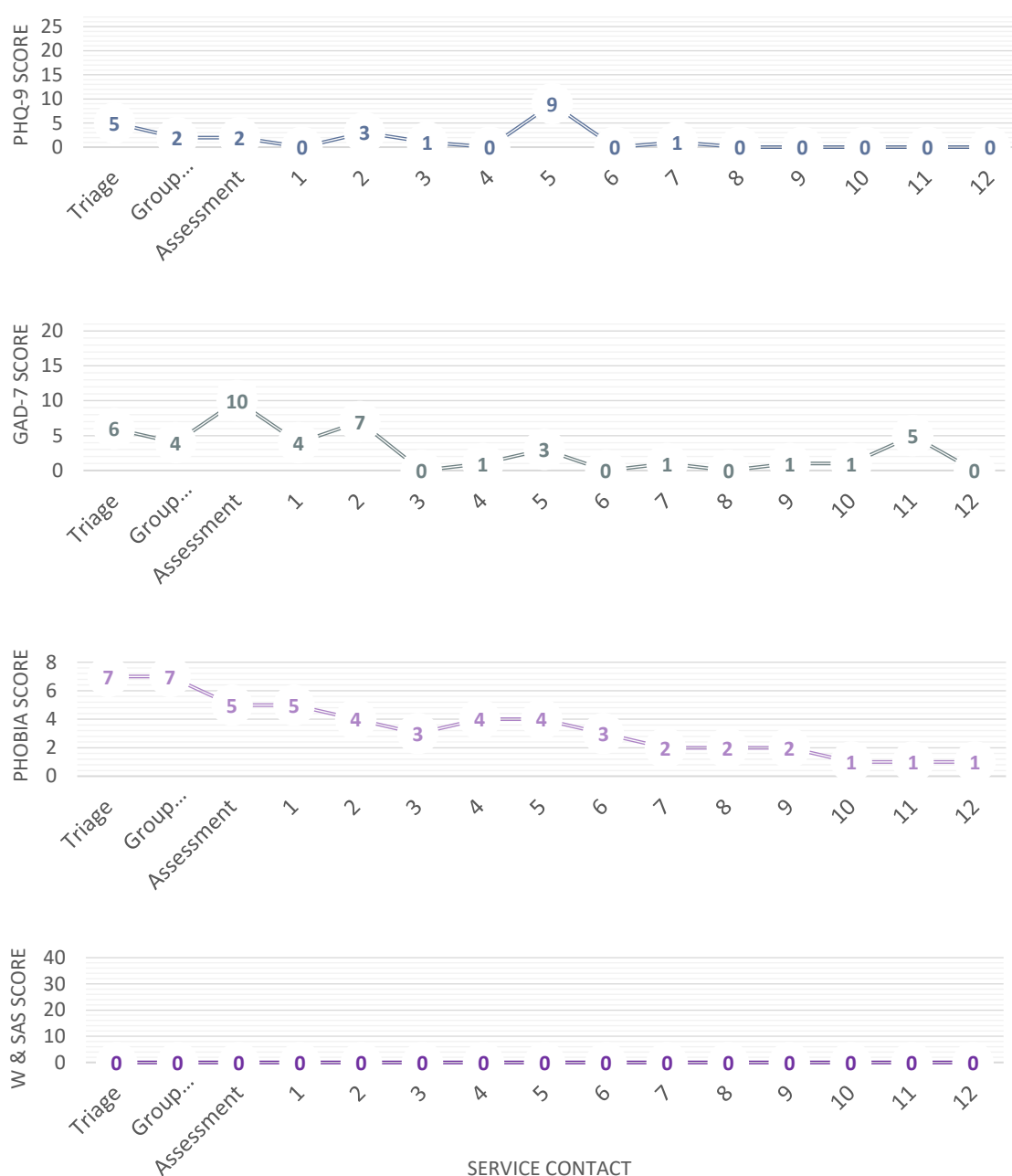
### 5.2. Outcome Measures

Please see Table 2 for a summary of Sophie’s outcome scores throughout treatment. Her PHQ-9 and GAD-7 scores did not reach clinical ‘caseness’ on triage, although her GAD-7 score of 10 at assessment did cross this cut-off. Positively, all outcome scores decreased over time, having minimal impact by discharge. This is with the exception of the W & SAS measure which remained constant at zero throughout treatment.

**Table 2: Sophie’s outcome scores at triage, assessment, and discharge for each outcome measure used, along with the measure’s range.**

	Range	Triage (19/06/15)	Assessment (19/11/15)	Discharge (25/02/16)
<i>PHQ-9</i>	0-27	5	2	0
<i>GAD-7</i>	0-21	6	10	0
<i>Specific Phobia</i>	0-8	7	5	1
<i>W &amp; SAS</i>	0-40	0	0	0

Please see Figure 2 for an outline of Sophie's outcome scores at each contact throughout therapy. Although four measures were used to monitor weekly outcomes (due to the service requirements), it was the specific phobia questionnaire that explicitly measured Sophie's presenting problem. Positively, Sophie's scored seven out of eight on this at the start of therapy, which gradually decreased by six points over sessions to one by discharge. Her PHQ-9/GAD-7 scores also generally decreased with each session, although Sophie was also able to attribute her increase in symptoms of low mood at session three and five to relationship difficulties (PHQ-9), and her increase in anxiety symptoms on occasional sessions to the exposure tasks (GAD-7).



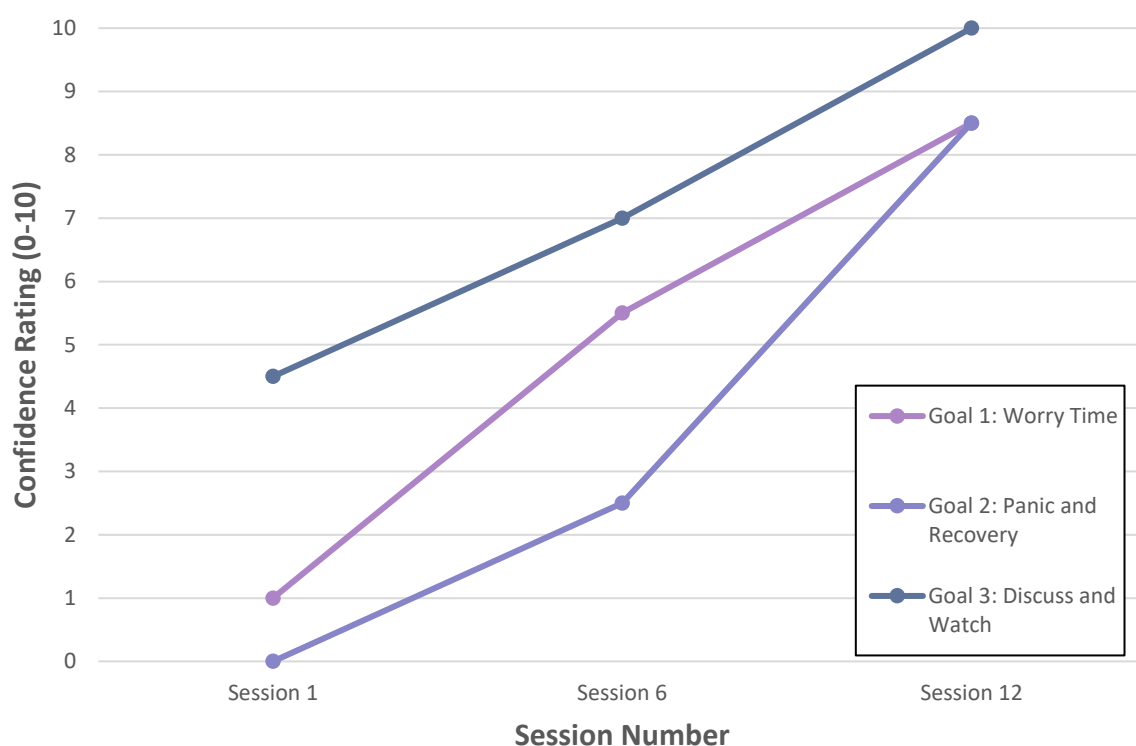
**Figure 2: Line graphs to show Sophie's outcome scores on each measure at each contact with the service.**

### 5.3. Goals

Sophie's goals were measured using a confidence rating scale (0-10) to indicate how much she felt she could achieve the goal now (0 indicating 'not at all', and 10 indicating 'definitely'). These were rated at the start, middle, and end of therapy. Her goals were to:

- (1) Spend less time worrying before a hospital/dental appointment;
- (2) Reduce the time spent physically panicking and recover from this faster;
- (3) Be able to watch TV programmes with needles and talk about needles calmly.

Figure 3 presents Sophie's ratings across the three time-points. Over therapy Sophie had made significant improvements on all three goals and by the end she subjectively considered herself as having achieved every one. Sophie was not concerned that goals two and three had not reached 10 out of 10, as she felt it was not realistic to experience absolutely no anxiety in these situations, and how this anxiety reaction is quite 'normal'.



**Figure 3: Line graph to show Sophie's confidence scores for each goal across therapy.**

## 6. DISCUSSION

---

### 6.1. Conclusions and Limitations

The current case presents an effective example of CBT for the treatment of a specific phobia of needles. As a result of using a collaborative and evidence-based approach, Sophie achieved a significant reduction in the associated psychological distress, and also within a limited number of 12 sessions. This is evident by the decrease in scores on all outcome measures, as well as the client's qualitative account that she achieved all three therapeutic goals, and the behavioural result of reaching the top of her fear ladder.

Sophie was encouraged to reflect on the factors that were maintaining the phobia (through the development of a formulation), and as Kirk and Rouf (2004) suggested, she also participated in exposure tasks that were designed to target the specific maladaptive thoughts that Sophie had expressed (e.g. that the needle will snap). All of which proved effective. This is likely to be a result of Sophie learning that she was overestimating the threat associated with the needle and underestimating her coping abilities. This is evident on her relapse prevention plan, where she wrote "the exposures have taught me that the object is not that bad, and it is to do with my thoughts" and "I can cope". These findings also support previous literature that exposure is an effective approach to treat specific phobia, and that facilitating multiple exposures, promoting cognitive change, and violating the client's expectations of the exposure are valuable strategies (Craske et al., 2014; Foa & Kozak 1986; Salkovskis 1991; Wolitzky-Taylor et al., 2008).

However, follow-up data was unable to be obtained and therefore it is unclear to whether the treatment effects have been sustained outside of treatment. Similarly, as the outcome measures are self-reported and reviewed by the therapist weekly, there may be bias in the results. For instance, they may be over-estimated due to a tendency for Sophie to respond in a favourable manner (social desirability bias). The outcomes used were also relatively brief, and it may have been beneficial to include a more in-depth measure to understand the specific factors relating to Sophie's phobia.

### 6.2. Reflections

Overall, I found the CBT model for treating specific phobia as very structured and effective. The main difficulties I faced was the feeling of disempowerment by the service due to the lack of resources for carrying out the exposure tasks. For instance, it was difficult to access actual needles; my contact attempts to the on-site phlebotomy clinic were not returned; and there was

no technology in the therapy rooms to present stimuli (e.g. pictures/videos). Nonetheless, the staff working at the service were extremely helpful, and one of the most memorable exposure tasks was when a colleague joined a session to discuss needles and volunteered to inject herself (as she had a nursing background and regularly did this for medical reasons).

Regarding the therapeutic relationship, I was conscious that Sophie was a full-time student and therefore may have had a tendency to view me like a teacher, i.e. in a position of power and as a dictator. I was particularly aware of this in the sessions that involved psycho-education, and for setting homework. I addressed this by being as collaborative as possible and taking a guided discovery approach. Sophie was also the first client I had ever provided with a full psychological intervention, and thus I immediately felt inadequate and unconfident in my capabilities. So, whilst there was possibly the expectation for me to be the 'teacher', I was avoidant of this and wanted to 'push the expertise' back to Sophie. Even so, this experience was extremely positive in terms of outcomes, which I was grateful for as it empowered me to be confident in my skills and work effectively with subsequent clients.

## 7. REFERENCES

---

- Alonso, J., Angermeyer, M. C., Bernert, S., Bruffaerts, R., Brugha T. S., Bryson, H., de Girolamo, G., de Graaf, R., Demyttenaere, K., Gasquet, I., Haro J. M., Katz, S. J., Kessler, R. C., Kovess, V., Lepine, J. P., Ormel, J., Polidori, G., Russo, L. J., Vilagut, G., Almansa, J., Arbabzadeh-Bouchez, S., Autonell, J., Bernal, M., Buist-Bouwman, M. A., Codony, M., Domingo-Salvany, A., Ferrer, M., Joo, S. S., Martinez-Alonso, M., Matschinger, H., Mazzi, F., Morgan, Z., Morosini, P., Palacin, C., Romera, B., Taub, N., Vollebergh, W. A., ESEMeD/MHEDEA 2000 Investigators, European Study of the Epidemiology of Mental Disorders (ESEMeD) Project. (2004). Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatrica Scandinavica Supplementum*, 109, 21- 27.
- Andrews, G. J. & Shaw, D. (2010). "So we started talking about a beach in Barbados": Visualization practices and needle phobia. *Social Science & Medicine*, 17(10), 1804-1810.
- Barlow, D. H., Moscovitch, D. A., & Micco, J. A. (2004). Psychotherapeutic interventions for phobias: A review. In M. Maj, H.S. Akiskal, J.J. Lopez-Ibor, & A. Okasha (Eds.), *Evidence and experience in psychiatry Volume 7: Phobias* (pp. 179–210). West Sussex: John Wiley & Sons Ltd.
- Choy, Y., Fyer, A. J., & Lipsitz, J. D. (2007). Treatment of specific phobia in adults. *Clinical Psychology Review*, 27, 266–286.
- Cisler, J. M., Lohr, J. M., Sawchuk, C. N., & Olatunji, B. O. (2010). Specific Phobia. In J. C. Thomas & M. Hersen (Eds.), *Handbook of Clinical Psychology Competencies* (pp. 697-722). New York: Springer.
- Coelho, C. M. & Purkis, H. (2009). The Origins of Specific Phobias: Influential Theories and Current Perspectives. *Review of General Psychology*, 13(4), 335–348.
- Craske, M., Treanor, M., Conway, C, Zbozinek, T. & Vervliet, B. (2014). Maximising exposure therapy: an inhibitory learning approach. *Behaviour Research and Therapy*, 58, 10-23.
- Curtis, G. C., Magee, W. J., Eaton, W. W., Wittchen, H. U., & Kessler, R. C. (1998). Specific fears and phobias. Epidemiology and classification. *The British Journal of Psychiatry*, 173, 212–217.

- Dudley, R. & Kuyken, W. (2006). Formulation in cognitive behavioural therapy. In L. Johnstone & R. Dallos (Eds.), *Formulation psychology and psychotherapy*. Routledge: Oxford.
- Foa, E.B. and Kozak, M.J. (1986). Emotional processing of fear: exposure to corrective information. *Psychological Bulletin*, 99, 20–35.
- Greenberger, D. & Padesky, C. A. (1995). *Mind over mood: A cognitive therapy treatment manual for clients*. New York, NY, US: Guilford Press.
- Hofmann, S. G. & Reinecke, M. A. (Eds.). (2010). *Cognitive-behavioral Therapy with Adults: A Guide to Empirically-informed Assessment and Intervention*. New York: Cambridge University Press.
- Hood, H. K. & Antony, M. M. (2012). Evidence-Based Assessment and Treatment of Specific Phobias in Adults. In T. E. Davis, T. H. Ollendick, & L. G. Öst (Eds.), *Intensive One-Session Treatment of Specific Phobias* (pp. 19-42). New York: Springer.
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62(6), 593–602.
- Kirk, J., & Rouf, K. (2004). Specific Phobias. In J. Bennett-Levy, G. Butler, M. Fennell, A. Hackmann, M. Mueller & D. Westbrook (Eds), *Oxford guide to behavioural experiments in cognitive therapy* (pp. 161-182). Oxford: Oxford University Press.
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606–13.
- Mower, O. H. (1939). Stimulus response theory of anxiety. *Psychological Review*, 46, 553-565.
- Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. M. (2002). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *The British Journal of Psychiatry*, 180, 461-464.
- Öst, L. G. & Sterner, U. (1986). Applied Tension: A specific behavioral method for treatment of blood phobia. *Behavior Research and Therapy*, 25(1), 25-29.
- Rachman, S. J. (1978). *Fear and courage (A Series of books in psychology)*. San Francisco: W.H.Freeman & Co Ltd.
- Salkovskis, P.M. (1991). The importance of behaviour in the maintenance of anxiety and panic: a cognitive account. *Behavioural Psychotherapy*, 19, 6–19.

- Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, 166(10), 1092–7.
- Willemsen, H., Chowdhury, U., & Briscall, L. (2002). Needle phobia in children: a discussion of aetiology and treatment options. *Clinical Child Psychology and Psychiatry*, 7(4), 609–619.
- Wolitzky-Taylor, K. B., Harowitz, J. D., Powers, M. B., & Telch, M. J. (2008). Psychological approaches in the treatment of specific phobias: A meta-analysis. *Clinical Psychology Review*, 28(6), 1021-1037.



## 8. APPENDICES

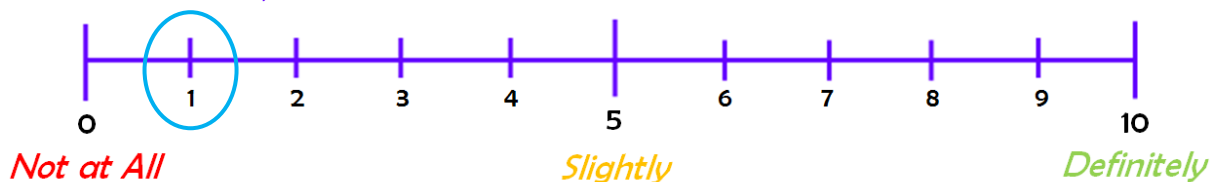
### 8.1. Goal-Setting Worksheet (used in Session 1)



#### **Goal 1:**

*"To reduce the amount of time I spend worrying about a hospital/dentist appointment, from days to hours"*

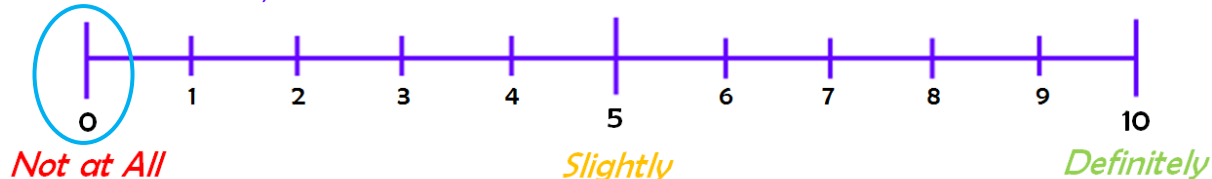
On the scale of 0 – 10, how confident am I that I can achieve this NOW?



#### **Goal 2:**

*"To reduce the time I spend physically panicking, so to spend less time feeling these sensations and be faster at recovering afterwards"*

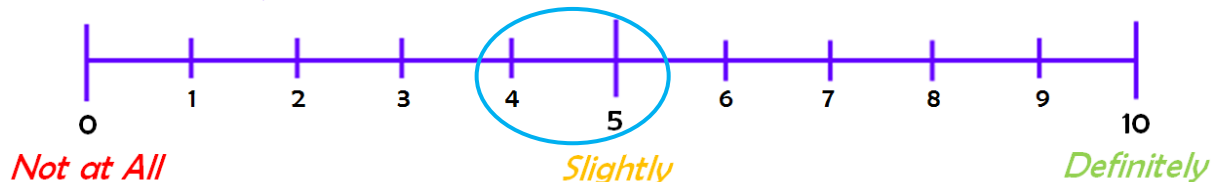
On the scale of 0 – 10, how confident am I that I can achieve this NOW?



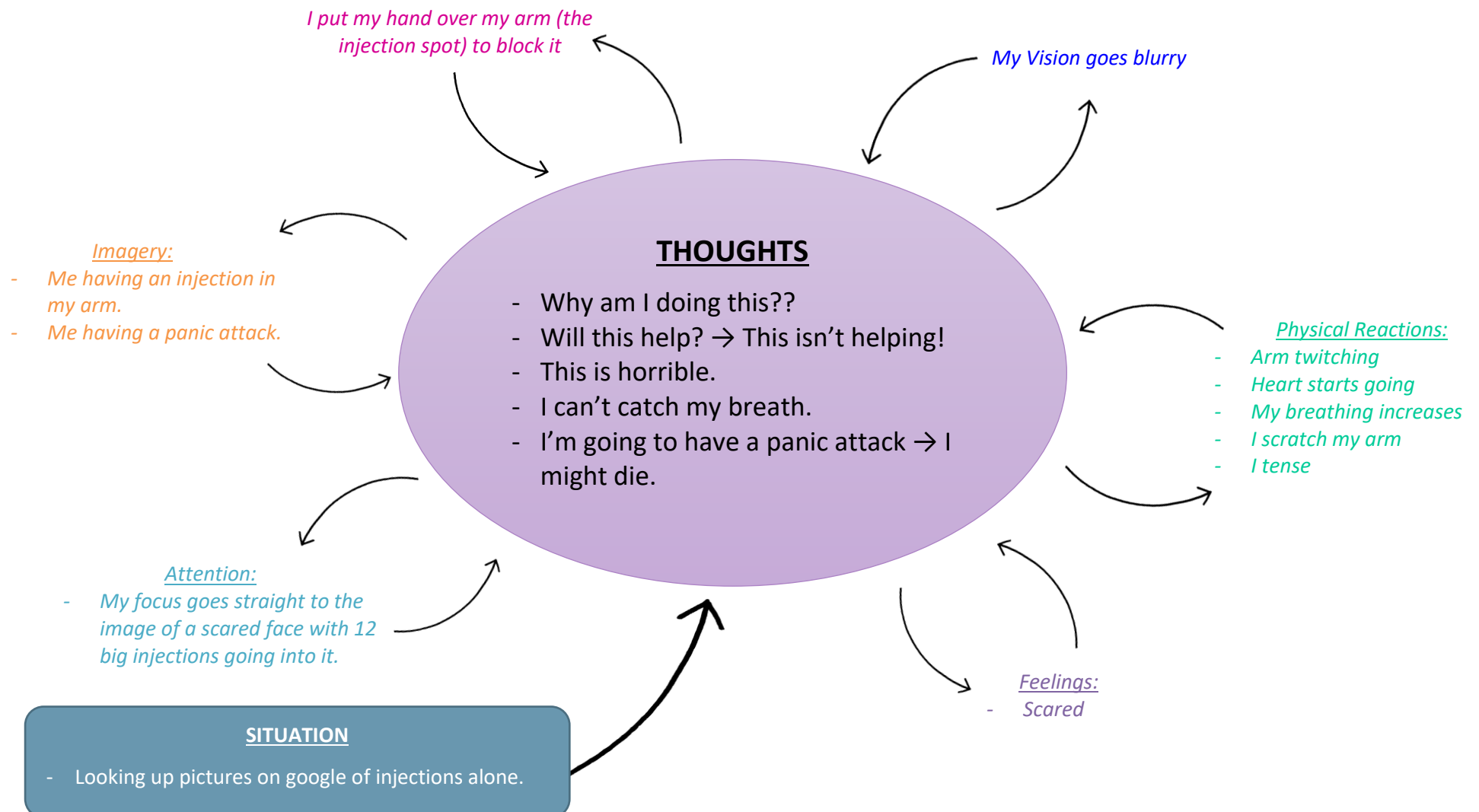
#### **Goal 3:**

*"To be able to talk about needles and to watch a TV programme with needles in (e.g. 24hrs in A&E) more calmly"*

On the scale of 0 – 10, how confident am I that I can achieve this NOW?



## 8.2. Vicious Flower Formulation (used in Session 2)



BY DOING THESE THINGS I AM KEEPING MY NEGATIVE THOUGHTS ABOUT INJECTIONS GOING, CREATING VICIOUS CYCLES. THIS INCREASES MY FEAR!

### 8.3. Fear Ladder Handout (used in Session 3)

## Making a Fear Ladder

*In the table below, brainstorm a list of situations / events / people that you avoid because of your fear. Write them in the table under the heading 'Things I avoid'. These can be in any order.*

*Once you have a list, rate how anxious they make you feel in the column 'Anxiety Rating'. This should be a number between 0-100; where 0 means no anxiety at all, and 100 means the most anxious you have ever felt.*

Things I Avoid	Anxiety Rating (0-100)
Having an actual injection	90
Having a blood test	85
Watching a medical programme	65
Looking at a medical display in a museum	20
Seeing the boy at school with the "Metal Leg Brace"	50
Watching a video of a bone marrow injection	80
Watching someone be tattooed	15
Looking at pictures of someone taking heroin	75
Having a conversation about needles	55
Talking to a doctor about travel vaccinations	60
Touching a fake needle	40
Holding a sewing needle	30
Looking at pictures of needles injecting people	70
Watching the scene from 'Saw' where a lady falls into a pit of needles	65
Looking at isolated pictures of needles	45
Having a real needle in the room in front of me	80

# My Fear Ladder

*Look at this list you just made of the situations which you avoid because of your fear.*

*Put the item you rated as the most anxiety provoking at the TOP of the ladder. Put the item you rated as the least anxiety provoking at the BOTTOM of the ladder.*

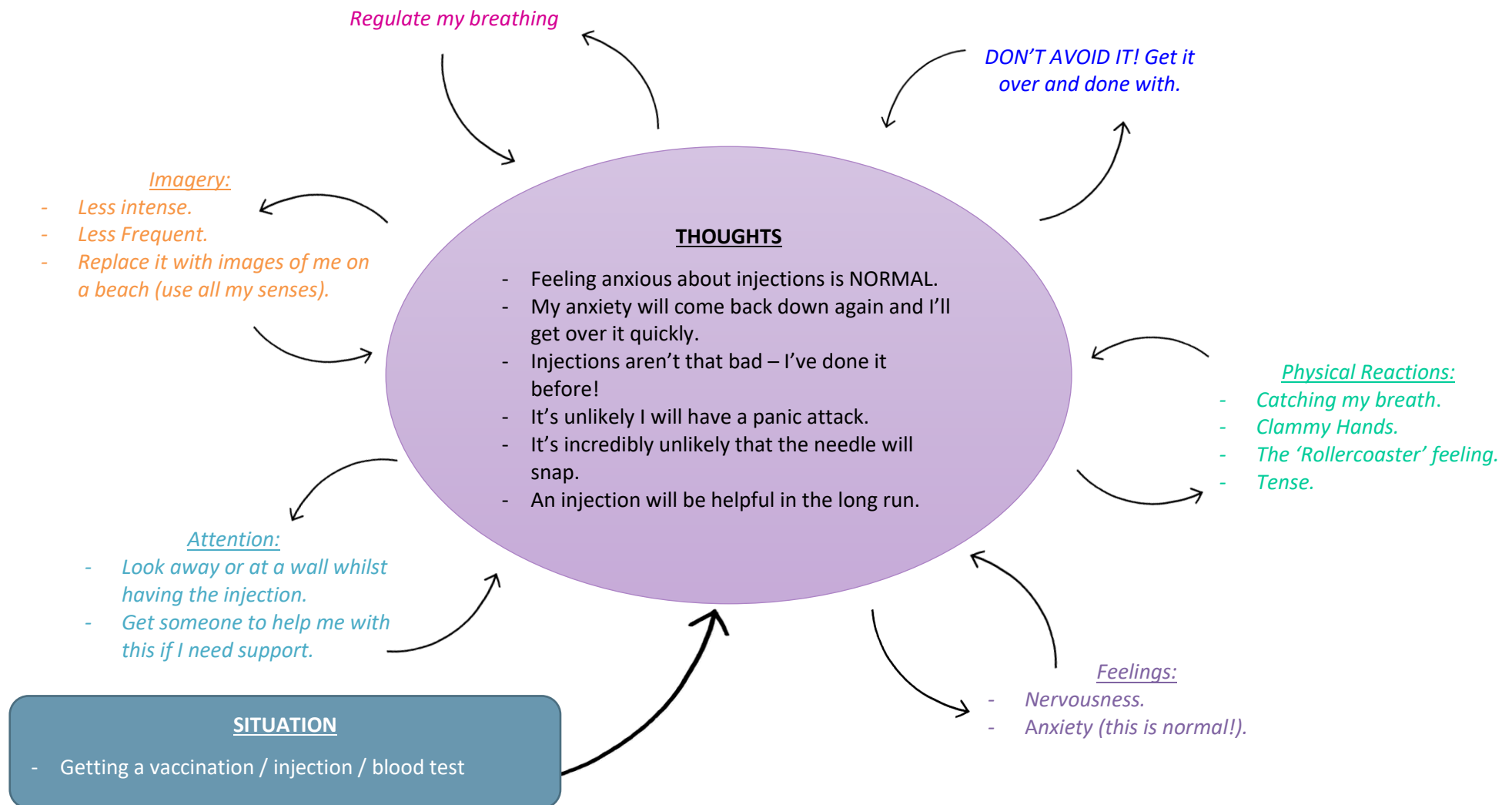
*Now fill in the other steps from high to low based on the ratings you gave them. If you rated some items the same, put them in an order which makes the most sense to you, so that your fear ladder steps move from your least feared situation to your most feared at the top.*

*It is okay if there are some blank steps.*



- 90 Having an actual injection
- 85 Having a blood test
- 80 Having a real needle in the room in front of me
- 80 Watching a video of a bone marrow injection
- 75 Looking at pictures of someone taking heroin
- 70 Looking at pictures of needles injecting people
- 65 Watching a medical programme
- 65 Watching scene from 'Saw' where a lady falls into a pit of needles
- 60 Talking to a doctor about travel vaccinations
- 55 Having a conversation about needles
- 50 Seeing the boy at school with the "Metal Leg Brace"
- 45 Looking at isolated pictures of needles
- 40 Touching a fake needle
- 30 Holding a sewing needle
- 20 Looking at a medical display in a museum
- 15 Watching someone be tattooed

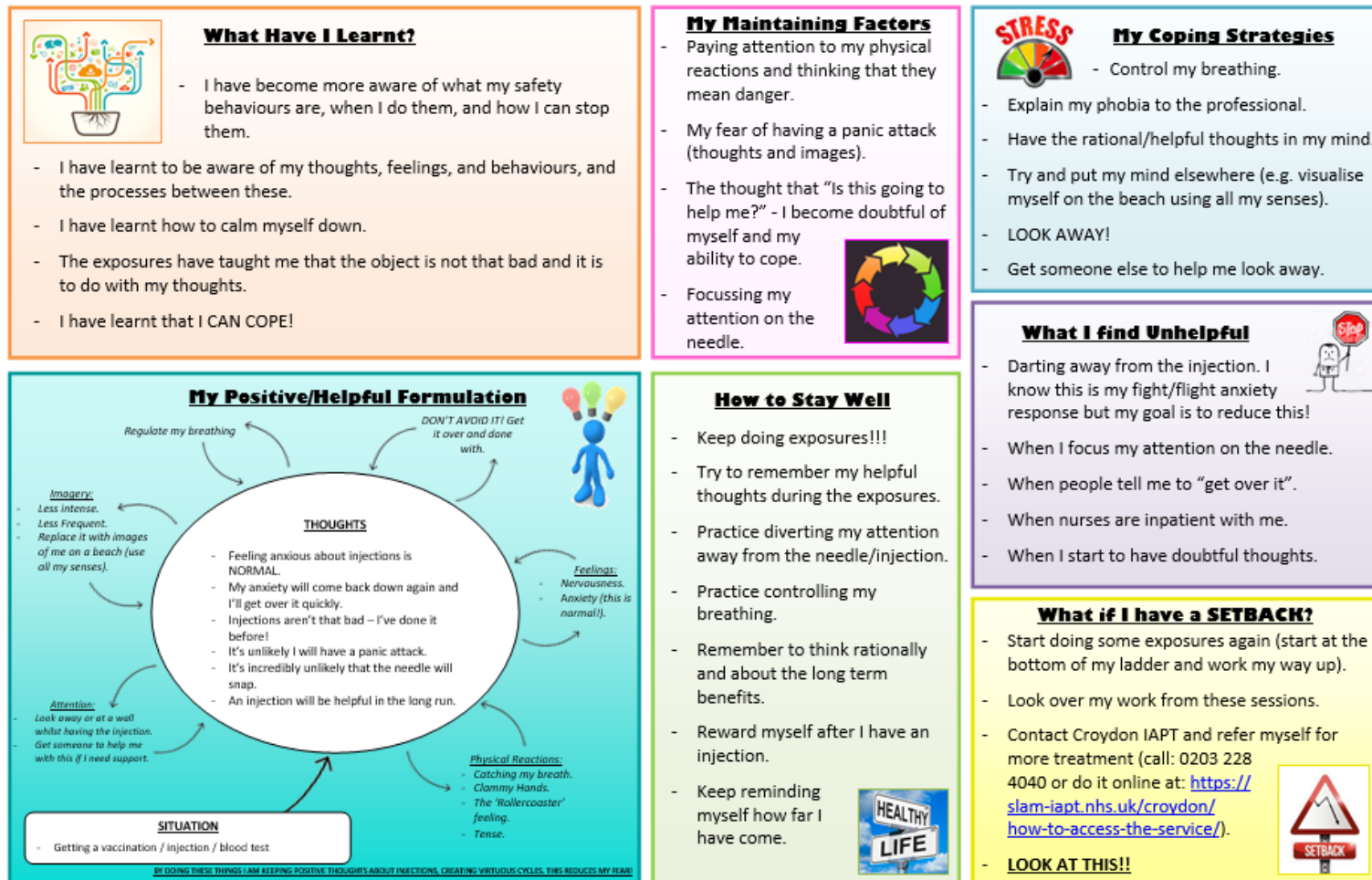
## 8.4. Virtuous Flower Formulation (used in Session 9)



BY DOING THESE THINGS I AM KEEPING POSITIVE THOUGHTS ABOUT INJECTIONS, CREATING VIRTUOUS CYCLES. THIS REDUCES MY FEAR!

## 8.5. Relapse Prevention Plan (used in Session 12)

# MY Relapse Prevention Plan



## 8.6. Reviewed Goals (used in Session 12)



### **Goal 1:**

*“To reduce the amount of time I spend worrying about a hospital/dentist appointment, from days to hours”*

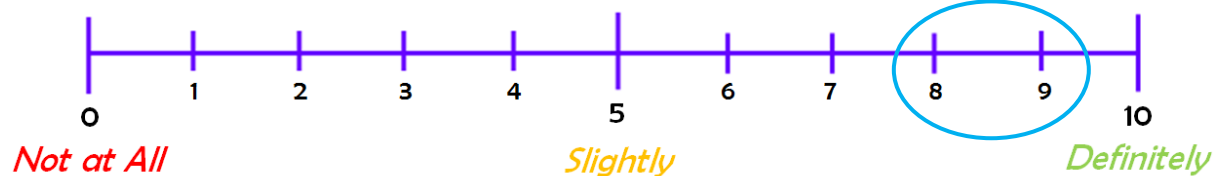
On the scale of 0 – 10, how confident am I that I can achieve this NOW?



### **Goal 2:**

*“To reduce the time I spend physically panicking, so to spend less time feeling these sensations and be faster at recovering afterwards”*

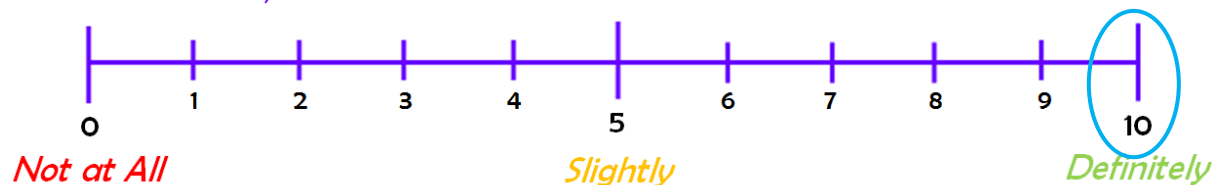
On the scale of 0 – 10, how confident am I that I can achieve this NOW?



### **Goal 3:**

*“To be able to talk about needles and to watch a TV programme with needles in (e.g. 24hrs in A&E) more calmly”*

On the scale of 0 – 10, how confident am I that I can achieve this NOW?



## CASE STUDY TWO

---

### **Overcoming the Anger:**

CBT with an 11-year-old boy

---

**Supervised by:** Dr Suraba Mahendiran



## ABSTRACT

---

“Tommy” was an 11-year-old boy previously known to CAMHS for diagnoses of ADHD and OCD. He was internally referred for CBT treatment for OCD, after reports of these symptoms becoming worse. However, at the assessment Tommy identified the primary problem as being anger-related, which was significantly interfering with his well-being, specifically at home. Taking a patient-centred approach, and with the aim of exploring the relationship between the OCD/ADHD and the anger at a later stage, Tommy was offered 16 sessions to help him to overcome the anger. However, Tommy and his family only attended eight of these sessions, allowing the following components to be covered: goal-setting; psycho-education of emotions/anger; formulation; anger monitoring; “cool down” techniques; and relapse prevention. Weekly outcome measures were used to assess Tommy’s well-being and clinical symptoms across treatment, which positively showed some improvement by the end of therapy. The limitations relating to this case and reflections are also discussed.

# CONTENTS

---

<b>1. LITERATURE REVIEW.....</b>	<b>36</b>
1.1. Anger .....	36
1.2. Psychological Models and Intervention .....	37
<b>2. CASE DESCRIPTION .....</b>	<b>38</b>
2.1. Background.....	38
2.2. Referral.....	38
2.3. Assessment.....	38
2.4. Outcome Measures .....	39
<b>3. FORMULATION.....</b>	<b>41</b>
<b>4. INTERVENTION .....</b>	<b>43</b>
4.1. Engagement .....	43
4.2. Session Structure.....	44
<b>5. OUTCOMES .....</b>	<b>47</b>
5.1. Behavioural Results .....	47
5.2. Outcome Measures .....	47
<b>6. DISCUSSION .....</b>	<b>50</b>
6.1. Conclusions and Limitations.....	50
6.2. Reflections.....	51
<b>7. REFERENCES.....</b>	<b>53</b>
<b>8. APPENDICES .....</b>	<b>55</b>
8.1. 'Magic Triangle' (Stallard, 2002) .....	55
8.2. Goal-Setting Worksheet (used in Session 1) .....	55
8.3. Anger Thermometer with Tommy's labels (used in session 3) .....	57
8.4. Reviewed Goals (reviewed in Session 8).....	58
8.5. Relapse Prevention Plan (used in Session 8).....	59

## 1. LITERATURE REVIEW

---

### 1.1. Anger

Anger can be considered one of the primary emotions that we experience, which is universal, innate, and automatic (Ekman, 1992). Anecdotally, anger-related problems are frequently the main reason for referrals to Child and Adolescents Mental Health Services (CAMHS) and are common in children who are referred for counselling/psychotherapy (Sukhodolsky, Kassinove, & Gorman, 2004). Despite this, and probably due to the increasing demands on services, these referrals are often not accepted due to anger being considered a behavioural, rather than mental health problem. However, anger is often a key feature of numerous diagnoses in young people, and can negatively interfere with various domains, including health, work, and social relationships (DiGiuseppe & Tafrate, 2007). DiGiuseppe and Tafrate (2007) observed that anger is not considered as debilitating as other emotions (e.g. anxiety/depression) and does not receive the same attention in research/clinical practice. They also argue that anger should be considered as an independent diagnostic category. Conversely, Novaco (2010) opposes this idea, particularly given the normality of anger; the lack of empirical grounds or coherent nosology; the likelihood that this would promote further pathologizing of individuals; and that by having a separate diagnostic category, anger would have to be removed as a feature of current disorders, although it is often best understood within the context of these.

Some view anger as an impulse-control problem, which can result in irritability and aggression, and shares common characteristics with neurological disorders such as Attention Deficit-Hyperactivity Disorder (ADHD) (DiGiuseppe & Tafrate, 2007). Kitchens, Rosén, and Braaten (1999) studied how anger and aggression differed between children who did and did not have an ADHD diagnosis. The authors found that children with ADHD self-reported experiencing more anger, and their mothers reported more aggressive behaviour compared to non-ADHD children. This anger was also expressed more often and in more dysfunctional ways. Anger has also been found to be associated with several other clinical disorders (including Post Traumatic Stress Disorder, mood, and personality disorders), where anger is conceptualised as being dysregulated and occurring without appropriate control, particularly regarding the activation, expression, and experience of anger (see Novaco, 2010 for review).

## 1.2. Psychological Models and Intervention

Several psychological theories have been proposed to conceptualise anger-related problems. Historically, the most influential models were built with a biological and evolutionary perspective, where anger was seen to have an adaptive purpose for survival. Psychologists in the nineteenth- and early-twentieth-century, such as Darwin, James, Freud, and Lorenz, also believed that anger was instinctual and had to be controlled, otherwise aggression would arise (DiGiuseppe & Tafrate, 2007). However, experiencing anger does not always lead to aggression, and so these theories were seen to be too simplistic and ignored the role of cognition.

Some of the subsequent theories included: the 'frustration-aggression hypothesis' (Dollard, Doob, Miller, Mower, & Sears, 1939), which proposed that aggression occurred when an individual wanted and expected a goal to be achieved; 'social learning theory' (Bandura, 1977), suggesting that aggressive behaviour is learnt through observing others and is reinforced; and 'excitation transfer theory' (Zillmann, 2003, cited in DiGiuseppe & Tafrate, 2007), proposing that aggressive behaviour occurs via three factors, that is: the activation of emotional arousal, reinforcement history, and higher order cognitive processing. Furthermore, Anderson and Bushman (2002) generated the 'general aggression model' to encompass all the robust theories available. Accordingly, aggression is considered the consequence of an interaction between situational (e.g. cues, incentives, biological states) and personal (e.g. beliefs, traits, and goals) factors. This impacts on the individuals' cognitive, emotional, and physiological response to stimuli, resulting in certain appraisals and decision making, and thus allowing aggressive behaviour to (or not) occur.

When it comes to the clinical treatment of anger, much of the literature evaluates interventions with a cognitive-behavioural therapy (CBT) approach, and so is considered the most predominant model (Beck & Fernandez, 1998). It is underpinned by learning theory and information processing, where anger is associated with a heightened physiological arousal and perception of threat – interventions therefore encourage the development of adaptive information processing, coping, interpersonal communication and problem-solving skills, as well as enhancing self-esteem (Beck & Fernandez, 1998; DiGiuseppe & Tafrate, 2007; Down, Willner, Watts, & Griffiths, 2011). In 2004, a meta-analysis evaluating the effectiveness of CBT interventions for anger-related problems in young people (Sukhodolsky et al., 2004) demonstrated that, from a sample of 21 published and 19 unpublished studies, CBT was considered an effective treatment approach (mean effect size was in the 'medium' range). The authors note that this is comparable to meta-analysis studying the effect of psychotherapy with anger-related problems in young people.

## 2. CASE DESCRIPTION

---

### 2.1. Background

'Tommy' (pseudonym) was an 11-year-old, white British male, who lived with his parents and four siblings (his mother was also pregnant). Tommy was the middle child, where his eldest sibling was 15-years-old, and youngest was 10 months. Tommy was previously known to CAMHS as he was receiving care for a diagnosis of ADHD. In 2014, he also attended six sessions of CBT for Obsessive Compulsive Disorder (OCD), however, this therapy discontinued after a lack of engagement.

### 2.2. Referral

In March 2015, Tommy was internally referred for more CBT after his mother reported (during an ADHD review) that the OCD symptoms had worsened. Tommy was on the waiting list for approximately one year before being seen for psychological treatment.

### 2.3. Assessment

Tommy attended a one-hour, face-to-face assessment session at the beginning of his treatment. His mother and two younger siblings were also present. The aim was to obtain a first-hand account of the current difficulties, and to understand how this was interfering with Tommy's life. Secondary to this, Tommy's understanding of the previous therapeutic work, and what aspects he found helpful/unhelpful were sought.

Regarding the previous CBT, Tommy reported that this had "really helped me" and "I don't think about it [OCD] as much anymore". Tommy's mother disagreed and reported that she had observed the OCD thoughts and behaviours regularly in Tommy's day-to-day functioning. Instead, Tommy felt the main problem was not being able to control the "out-bursts" and anger. Despite exploring whether this was associated with thoughts consistent with OCD, Tommy denied this, and stated that they were usually triggered when he believed people were "being rude to me" (e.g. when his siblings took his toys). Tommy was unable to identify any thoughts at these times but noticed that behavioural consequences were acting aggressively towards objects and "saying nasty things" (usually to his mother). Emotional consequences were feeling "upset" for being hurtful. Tommy's mother did report however, that this was only a problem at a home, as Tommy was able to control the anger at school. This behaviour had been increasing over the previous year, and Tommy reported that this was interfering with his relationships with

his family members, and resulted in unwanted negative consequences, such as not being allowed to see his friends.

Anger is not typically treated within a tier-three CAMHS service as it is not perceived to be a mental health problem, however at this time, the influence of the OCD was still unclear, particularly considering the reports from Tommy's mother and CAMHS referral. We therefore decided to begin therapy with a focus on anger (Tommy's main concern and goal) but agreed to revisit the OCD at a later stage and explore their relationship together. In a child-friendly manner, confidentiality and the boundaries of therapy were explained, and no significant risk issues were identified. The risk of damaging property was explored, and this was currently being well-managed by his mother. The expectations and structure of therapy were discussed, as well as the basic theoretical components of CBT.

## 2.4. Outcome Measures

Outcome measures were used to measure the effectiveness of treatment, please see Table 3 for a comprehensive list. Two pre-/post-therapy measures were utilised, one assessed general clinical symptoms and screened for anxiety and mood disorders (RCADS), and the other was specific to OCD. Due to there being discrepancies between Tommy's and his mother's reports about the OCD, it was felt important to still measure this, to test this hypothesis and consider its relationship with the anger. Tommy understood this rationale and consented.

Weekly outcome measures were used to track changes within therapy, such as Tommy's general well-being and functioning, and the OCD symptoms. An anger measure was introduced later during therapy, because initially, no appropriate measure was available within the service (as anger-related problems were not typically seen, only in the context of a mental disorder). As therapy had already commenced, a pre-/post-outcome measure for anger was also not feasible. All measures used are recommended by the 'Membership of Outcomes and Evaluation Task and Finish Group' of the 'Children and Young Peoples' Increasing Access to Psychological Therapies' service (CYP IAPT OEG; 2011).

**Table 3: A list and description of each outcome measure used in therapy.**

<u>Outcome Measure</u>	<u>Description</u>
<b>PRE- / POST- MEASURES</b>	
<b>The Revised Child Anxiety and Depression Scale (RCADS)</b> (Chorpita, Yim, Moffitt, Umemoto, & Francis, 2000)	A 47-item self-report questionnaire which assesses symptoms of anxiety and depression consistent to diagnoses from the DSM-IV (major mood disorder, generalised anxiety disorder, panic disorder, OCD, separation anxiety disorder, social phobia). The RCADS has been shown to have good reliability and validity.
<b>The Child Obsessive-Compulsive Inventory – Revised (ChOCI-R)</b> (Uher, Heyman, Turner & Shafran, 2008)	A 32-item self-report questionnaire which assesses the content and severity of OCD symptoms. This measure has been shown to have good internal consistency, convergent validity, and divergent validity.
<b>WEEKLY MEASURES</b>	
<b>Child Outcome Rating Scale (CORS)</b> (Duncan, Sparks, Miller, Bohanske & Claud, 2006)	A 4-item self-report measure for children (6-12 years), assessing four areas of functioning/well-being (individually, interpersonally, socially, and overall), which has been shown to change as a result of therapeutic intervention. The CORS has demonstrated good validity, reliability, and feasibility.
<b>‘OCD Tracker’ from the RCADS</b> (Uher et al., 2008)	This brief self-report measure consists of the six questions within the RCADS OCD subscale.
<b>Goal Progress</b>	This is not a standardised measure but requires the child to rate, on a scale of 0-10 how much they feel they have progressed towards achieving their goal(s) (‘0’ meaning not at all and ‘10’ mean completely).
<b>‘Out of Control Behaviour Tracker’ from the Me and My School Questionnaire (M&amp;MS)</b> (Department for Education, 2010)	A 7-item self-report measure corresponding with those from the behavioural difficulties subscale of the M&MS. The M&MS was originally designed to evaluate the ‘Targeted Mental Health in Schools’ (TaMHS) government programme, which supported schools to deliver interventions according to their local mental health needs. The measure also shows good construct validity.

### 3. FORMULATION

A cross-sectional formulation was hypothesised following assessment and reviewed throughout therapy. The ‘hot-cross bun’ model by Greenberger and Padesky (1995) was utilised, as it was felt this was the most child-friendly and meaningful for Tommy’s age. Although Tommy was presented with a simplified version, namely the “magic triangle”, taken from Stallard’s (2002) “Think Good, Feel Good” book (Appendix 8.1). This does not include the physiological reactions to anger, but this was covered in separate sessions.

Please see Figure 4 below for a diagram of Tommy’s formulation. Tommy described the problem as “anger” and feeling unable to “control” this mood. When the anger was triggered, usually in situations which Tommy perceived to be “unfair” or “not right”, Tommy would experience thoughts such as “I want to hit you” and “this is unfair”. In his body, Tommy reported experiencing a “fire” in his belly and noticed that he would become hot and sweaty. Emotionally, he described feeling angry and annoyed, and he would subsequently engage in behaviours such as shouting, “back chatting” to his mum, and occasionally punching objects.

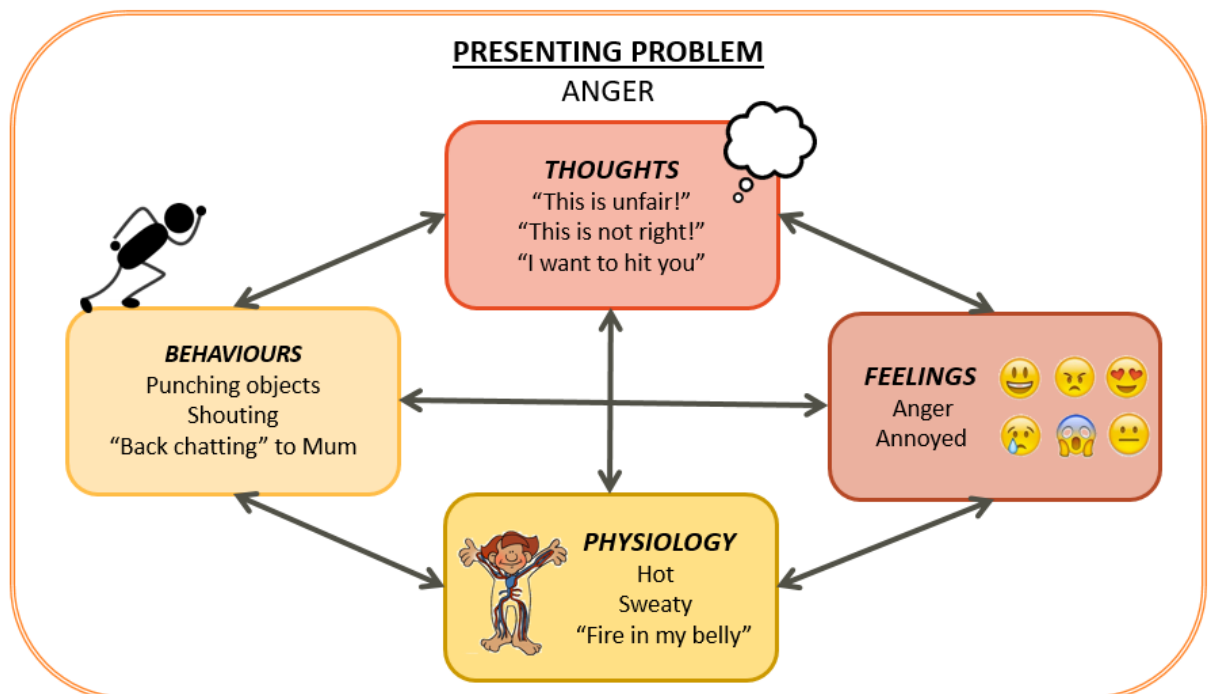


Figure 4: A cross-sectional formulation of Tommy’s presenting problem.



In addition to Tommy's child-friendly formulation, a more extensive formulation was hypothesised (but not shared) to understand the function and development of the anger in more detail. The framework used for this was the 'Five Ps' model by Dudley & Kuyken, 2006. Please see Table 4 below:

**Table 4: A hypothesised formulation for the Anger using the Five Ps Model (Dudley & Kuyken, 2006)**

<u>Description</u>	
<b>1. Presenting</b>	Anger
<b>2. Predisposing</b>	Tommy had a pre-existing diagnosis of ADHD, a neurodevelopmental disorder characterised by impulsivity. Therefore, Tommy would frequently act without thinking and may have lacked adequate skills to manage difficult emotions, subsequently increasing the likelihood of him becoming engaged in aggressive behaviour when the anger was triggered. Similarly, Tommy had a diagnosis of OCD, whereby he would frequently experience high levels of anxiety. One consequence of anxiety can be irritability, which may additionally have led to the experience of anger.
<b>3. Precipitating</b>	During the time before the onset of the anger, Tommy's mother would have just given birth to her youngest child. This was likely to have changed the family dynamics at home, which Tommy may have found difficult considering how his parents' attention would now need to be shared between five children. This may have been further exacerbated by his mother's current pregnancy of her sixth child. Tommy was also of an age where he was transitioning between schools, a difficult and crucial time for any child, where Tommy may have experienced additional needs for his parents' support.
<b>4. Perpetuating</b>	Tommy's "out bursts" and anger were likely to have been positively reinforced within his home environment by his parents. For instance, by acting aggressively it is possible that Tommy would then obtain the desired attention from his parents (albeit negative attention), who would subsequently interact with Tommy to implement the appropriate consequences to such aggressive behaviour. As a result, Tommy would have learnt that this behaviour is a way to meet his needs of attracting his parents' attention within the busy household.
<b>5. Protective</b>	Positively, Tommy could manage the anger whilst at school, suggesting that he had the skills to overcome the anger (and further implicating the role of the familial system). Additionally, during the sessions he would engage well and was assertive, for instance requesting to work on the anger and not OCD.

## 4. INTERVENTION

---

There are currently no guidelines available for the treatment of anger in young people, however CBT is the most commonly used approach in the literature (Beck & Fernandez, 1998). This model was also flexible, meaning it could be adapted to meet Tommy's current cognitive ability. Therefore, CBT was the adopted approach for this case. Further adaptations were also considered regarding the comorbid ADHD diagnosis; however, this was well-managed and so little action was required. Short, varying activities were utilised where possible, and Tommy often took a toilet-break towards the middle-end of sessions. These were sufficient in maintaining his attention throughout therapy.

Treatment consisted of eight one-hour weekly sessions. Each session followed a CBT framework of: setting the agenda, bridging from the previous session, reviewing any take-home tasks, carrying out collaboratively set activities, and setting new take-home tasks. Time for Tommy/his mother to provide feedback and confirm the next appointment was also available.

### 4.1. Engagement

In total, Tommy was offered 16 therapy sessions over a four-month period. However, Tommy (and his mother) did not attend six of these and cancelled two. This resulted in sessions being inconsistently spaced and content being missed (e.g. exploring the anger in the context of the ADHD/OCD). Due to limited childcare and Tommy not wishing to be alone, his mother and at least one sibling would be present in most sessions. This would become distracting at times, but Tommy managed this well and would re-focus when prompted by the therapist. This also meant that Tommy's mother was less engaged in the sessions as a co-therapist, as she was required to attend to the younger siblings' needs. Within sessions, Tommy always engaged well and asked appropriate questions when he did not understand. He appeared to be actively listening to the content, as evident by his ability to provide appropriate feedback. He put in effort to each task but would require some prompting to recall the contents of previous sessions. This was possibly due to the sporadic nature of sessions, and Tommy never completing take-home tasks between sessions – despite numerous conversations/negotiations around this.

## 4.2. Session Structure

### **Session 1: Assessment, Engagement and Goal-Setting**

The first session was primarily used to assess the presenting problem (as discussed previously in section 2.3). However, to begin with, time was spent getting to know Tommy to help improve his confidence in talking with a new therapist and make him feel more comfortable. As Tommy described the anger, a 'feelings wheel' was used to help Tommy think about the anger in relation to other emotions and to set 'SMART' therapy goals (Specific, Measurable, Achievable, Realistic and Time-limited). Please see Appendix 8.2 for Tommy's Goal-Setting Worksheet. This discussion also involved normalising anger, and briefly highlighting the importance and helpful functions of anger.

### **Session 2: Understanding Emotions (Psycho-education)**

This session focused on understanding what different emotions look like in other people (facial expressions, body language, and behaviours), and how they feel inside ourselves (physiological reactions). Due to the limits of time, only anger, happiness, and worry were focused on. This was a drawing task and Tommy was provided with templates of bodies and faces. Tommy also wrote a thought into a bubble for each emotion, a concept he appeared to grasp with support, and he provided appropriate examples.

### **Session 3: Anger Thermometer and the 'Magic Triangle' (Psycho-education and Formulation)**

During the session's "check-in" time, Tommy reported having a "bad week" and experiencing lots of anger. This was subsequently explored, and he recognised that the trigger was his friends "teasing" him on his school trip. The concept of anger triggers was discussed in more detail, and Tommy was introduced to the 'anger thermometer' (Appendix 8.3), illustrating the different intensities of anger. Tommy named each level on the thermometer and noticed the differences between these and how they feel. He then rated the anger he experienced relating to this event, as well as other events recalled from previous sessions. Psycho-education was also provided around the accumulation of anger and the importance of "cooling down".

The "magic triangle" was also introduced (Appendix 8.1), demonstrating the link between thoughts, feelings, and behaviours, whilst working through the anger event that he had brought to session. Additional activities were undertaken to help Tommy develop this understanding,

such as labelling a list of statements as thoughts/feelings/behaviours and completing empty thought bubbles on cartoon pictures.

#### **Sessions 4-7: “Cooling Down” Techniques**

The concept of the anger thermometer helped provide a rationale for learning techniques to “cool” the anger down. We considered how anger changes before, during, and after a triggered event, and drew a graph to illustrate this. Tommy recognised that the higher up the thermometer the harder it was for him to act and “cool down”, emphasising the importance of recognising the anger as early as possible. Tommy was taught several “cooling down” techniques across sessions. These included:

- *Two relaxation strategies*: controlled breathing (“4-2-6”) and progressive muscle relaxation (“the lemons”).
- *Identifying Warning signs*: listing these, and then developing a code word (“bananas”) for when Tommy or his mother noticed these.
- *Time Out*: leaving the anger situation to a calmer environment.
- *Anger Time*: an allocated slot every day where Tommy could discuss with his mother any events that had made him angry. He would note these down on paper throughout the day as they occurred.
- *Calming thoughts*: alternative thoughts he could think of instead of angry ones (e.g. “this won’t last forever” and “they haven’t got as many crew credits as me!!”).
- *My happy place*: Imagery task using his senses (this was a car garage).
- *Chill-out box*: a container with several calming stimuli (e.g. sock stress-balls, pictures of his favourite cars, and his school certificates).

Throughout these sessions, the “magic triangle” was also revisited regularly, whenever Tommy described an event from the week that had triggered the anger. This helped him to become more aware of the anger thoughts, feelings, and behaviours, although he still required some support with accessing the anger thoughts.

#### **Session 8: Review and Relapse Prevention Plan**

The final session was designed to review the work/progress Tommy had made over therapy, and to develop a plan integrating this together, giving Tommy clear directions about what he could do should the anger return in future. Tommy was presented with a folder containing all the work

he had completed. This was reviewed and discussed, as were his goals for therapy and the progress made towards these (please see section 5.3 below and Appendix 8.2/8.4 for more details). Tommy reported that for him to continue with this progress at home, he planned to refer to this folder regularly and “use all of the techniques and not just one”.

The plan was devised using a poster-design task, with the anger thermometer as a template. Tommy arranged puzzle pieces to produce the pictorial cue that represented each technique, and then ordered these on his poster according to which method he thought would be the most helpful to start cooling the anger down. These also linked with the intensity of anger on the thermometer, as the stronger the anger was, the more methods he would be required to recruit. Once completed, Tommy presented the poster to his mother, and a location for this and the folder was agreed. Please see Appendix 8.5 for a copy of Tommy’s anger thermometer and plan.

## 5. OUTCOMES

---

### 5.1. Behavioural Results

Tommy began therapy reporting not being able to control the anger that he experienced and described unhelpful strategies to manage this. Over the course of treatment, Tommy became more confident in talking about the anger, and more aware of the impact this had on his thinking, feelings, and actions, as evident by his references to the “magic triangle” and bringing anger examples to sessions. However, there were no overt behavioural changes that indicated that the anger had reduced in frequency/intensity. Due to the lack of engagement outside of sessions, coupled with poor attendance, it was unclear how generalisable and effective the skills were to Tommy’s home environment. Nonetheless, Tommy reported in the final session that “I am now 15% angry, instead of 45%”. He also described a positive experience of therapy, where he had “learnt that I can control my anger, the anger can’t control me”. Tommy’s mother was unable to attend the final sessions due to child care for Tommy’s other siblings. Therefore, her account on what changes she had observed at home was unable to be attained.

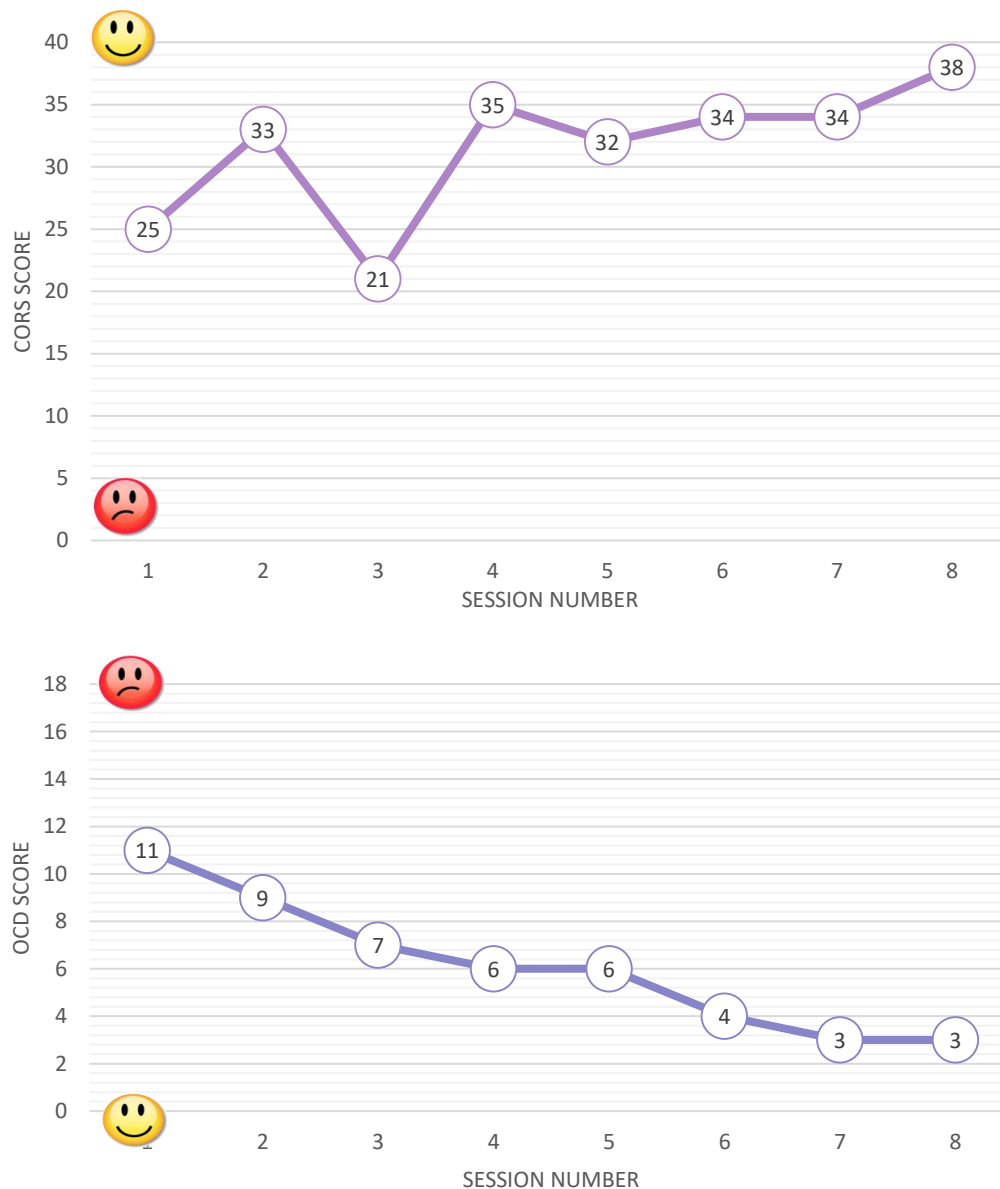
### 5.2. Outcome Measures

#### **Pre-/Post-Outcomes:**

Due to the take-home tasks not being completed, pre-intervention data was unable to be attained, and so no comparisons of Tommy’s well-being could be made before and after treatment. A post-RCADS was completed however, which revealed that Tommy was considered as “unlikely” to have any of the six DSM-IV disorders, including OCD.

#### **Weekly Outcomes:**

Tommy’s CORS and OCD scores across treatment can be found in Figure 5. Positively, his general well-being (CORS) increased throughout treatment by 13 points. There was one significant decline in scores during session three, which correlates with the “bad week” he described having in session. Similarly, a positive effect on the OCD symptoms were reported, despite this not being a primary focus of therapy. This reduced by eight points, from 11 to three by the end of therapy. The anger-related outcomes were considered insufficient to evaluate change across therapy as they were only obtained in sessions 6-8, nonetheless, by the end of therapy this was within the “normal” range.

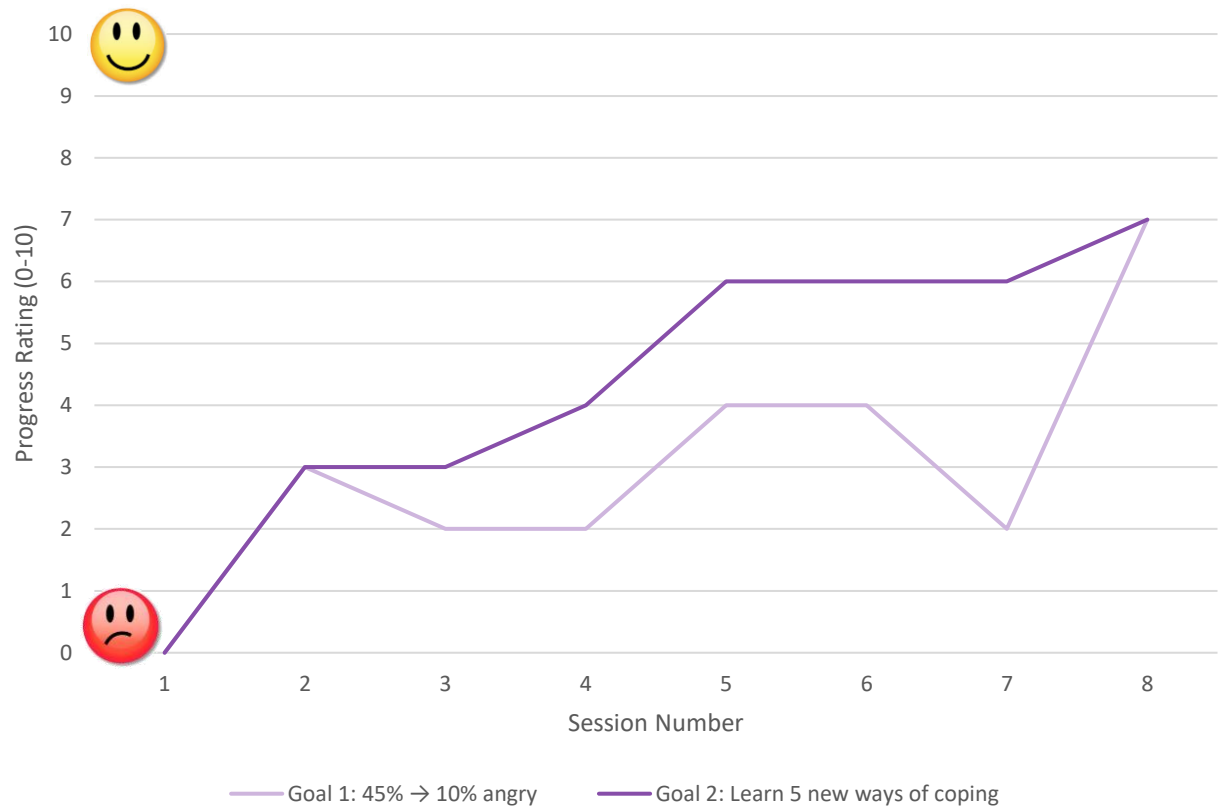


**Figure 5: Line graphs to show Tommy's outcome scores on each measure at each contact with the service.**

Tommy also rated his progression towards the goals set at the beginning of therapy, which were:

- (1) "To go from being 45% angry to 10% angry."
- (2) "To go from knowing one unhelpful way of coping with anger (taking it out on Mum) to learning five new helpful ways of coping."

Figure 6 below presents Tommy's progression ratings across each session. For both goals, Tommy began therapy rating these at zero, which increased to seven by the end of therapy. For goal two, the inclination from zero to seven out of ten was steady across therapy, whereas there were more fluctuations in ratings regarding goal one – especially from week seven to eight which increased by five.



**Figure 6: Line graph to show Tommy's subjective goal progress scores for each goal across therapy.**



## 6. DISCUSSION

---

### 6.1. Conclusions and Limitations

This case presented an example of adapted-CBT for the treatment of anger with a child. Using an evidenced-based approach and being patient-centred resulted in improvements in Tommy's well-being, and considerable progress was made towards his therapeutic goals, as evident by the weekly outcome scores and Tommy's qualitative account. Within their meta-analysis, Sukhodolsky and colleagues (2004) found that varying effects were obtained when treatments were categorised according to the target of therapy (i.e. skill development, affective education, problem solving, or eclectic treatments), with skills training and eclectic interventions proving most efficient. The authors interpreted these findings to suggest that treatments are more effective when they are "more behavioural" and teach actual behaviours rather than only focusing on internal, cognitive constructs. Further, they also noted that treatment duration (ranging from 2-30 hours) had no significant influence on the findings. Therefore, even though the breadth of treatment for this case was not as long as initially hoped, this should not have had a significant effect on the intervention's outcomes, and rather it is the intervention's content that is important. Due to Tommy's age and time available, the behavioural skills were emphasised more than the cognitive ones, which can possibly explain the positive remarks Tommy made about therapy.

A secondary, unexpected, therapeutic gain was an improvement in OCD symptoms, which was the original reason for referral but was unable to be explicitly targeted in treatment due to time constraints. There was a hypothesis that this could have been a trigger for the anger, or related to the symptoms in another way, however this was unable to be explored. Regardless, Tommy did not support this account, and evidence has shown that there are no differences in anger between OCD patients and controls, and that any anger expression from these individuals is usually not a component of the disorder but a reflection of general distress (Whiteside & Abramowitz, 2005). Therefore, this could suggest that the OCD symptoms reduced through the development of coping strategies, an increase in emotional awareness, and a general increase in well-being.

However behaviourally, there was little evidence reflecting the same subjective changes that Tommy described, which may be explained by several factors, such as poor attendance and a lack of generalising the skills learnt in therapy to his personal life. However, due to the lack of follow-up data, the sustainability of the treatment's effects is unclear. Additionally, Tommy's

age could have played a role, as Down and colleagues (2011) found that treatment outcomes from a CBT group-intervention for anger management was correlated to age. That is, younger participants were less engaged and motivated, and so obtained poorer treatment outcomes. Specifically, these participants were observed to: be dismissive of techniques; be more likely to attribute others as being responsible for the anger (believing others should change first); and fear of peer ridicule or their own reactions if they disclosed personal material and emotions. Although this evaluation was of group-therapy, some features may still reside in individual therapy, especially for Tommy who was often speaking in front of the therapist, his mother, and siblings.

Moreover, this case is limited by the quality of the outcome measures used. For instance, they were self-report measures which were reviewed in-session with the therapist, allowing the responses to be influenced by a social desirability bias and therefore may be an over-estimation of the true therapeutic effects. These were also relatively brief, where a more in-depth evaluation may have provided richer information and allowed for a greater understanding of the therapeutic changes. Even more so for the measure of anger, which unfortunately was absent for most of the therapy. If a similar piece of work was encountered in future, it may prove helpful to use the 'Beck Anger Inventory' for example, although such measures can incur a cost to the service to use.

## 6.2. Reflections

This CBT model was very useful in developing strategies for Tommy to overcome the anger, and it was reassuring to see that working in a patient-centred manner can have positive effects on secondary clinical symptoms. I believe one of the most influential components of the therapy was externalising the anger, which appeared to give Tommy a sense of empowerment and know that he was not the problem, the anger was. Although a narrative therapy technique, it demonstrated to me the value of combining techniques from various models to compliment the CBT work, and be a more holistic therapist. On hindsight, I wondered whether a systemic approach may have also been more influential. I attempted to encourage Tommy's mother to be a co-therapist and to understand the system at home, however, due to Tommy's younger siblings being in the room this was too difficult, and the mother was often distracted by the needs of these siblings. I felt as though this may also reflect Tommy's home environment, which made me re-evaluate the function of the anger (i.e. a reinforced strategy to obtain attention).

Therefore, a family therapy referral may have been more appropriate and effective at targeting the anger.

One of the main frustrations I had was the experience of poor engagement, meaning that time did not permit me to cover topics that I felt could have been of value, such as the family system and comorbidities. This felt like the biggest hurdle and was something I considered when ending therapy. For instance, I could have continued the work as a carry-over case or referred him to another therapist in the service. However, despite highlighting the importance of engagement with the client and his mother and witnessing no change in attendance and work outside of sessions, I felt sure that this pattern would continue regardless, and thus not add any significant benefits for the client. This also made me question the utility of DNA policies, which some services strictly implement. This case has made me feel more comfortable sitting with feelings of dissatisfaction at the progression made and helping the client and their family also take responsibility for the work, rather than me as the therapist taking full responsibility.

## 7. REFERENCES

---

- Anderson, C. A., & Bushman, B. J. (2002). Human Aggression. *Annual Review of Psychology*, 53, 27-51.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavior change. *Psychological Review*, 84, 191-215.
- Beck, R., & Fernandez, E. (1998). Cognitive-behavioral therapy in the treatment of anger: A meta-analysis. *Cognitive therapy and research*, 22, 63-74.
- Children and Young Peoples' Increasing Access to Psychological Therapies: Membership of Outcomes and Evaluation Task and Finish Group (CYP IAPT OEG). (2011). *Children & Young People's IAPT Routine Outcome Monitoring: Briefing Note*. Retrieved from <http://www.cypiapt.org/site-files/rom-dec11-03.pdf>.
- Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L. A., & Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A revised child anxiety and depression scale. *Behaviour Research & Therapy*, 38, 835-855.
- Department for Education. (2010). *Me and My School: Findings from the National Evaluation of Targeted Mental Health in Schools 2008-2011* (reference: DFE-RR177). Retrieved from [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/184060/DFE-RR177.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/184060/DFE-RR177.pdf).
- DiGiuseppe, R., & Tafrate, R. C. (2007). *Understanding Anger Disorders*. New York, New York: Oxford University Press.
- Dollard, J., Doob, L. W., Miller, N. E., Mowrer, O. H. & Sears, R. R. (1939). *Frustration and Aggression*. New Haven, Connecticut: Yale University Press.
- Down, R., Willner, P., Watts, L, & Griffiths, J. (2011). Anger management groups for adolescents: A mixed-methods study of efficacy and treatment preferences. *Clinical Child Psychology and Psychiatry*, 16, 33-52.
- Dudley, R. & Kuyken, W. (2006). Formulation in cognitive behavioural therapy. In L. Johnstone & R. Dallos (Eds.), *Formulation psychology and psychotherapy*. Routledge: Oxford.

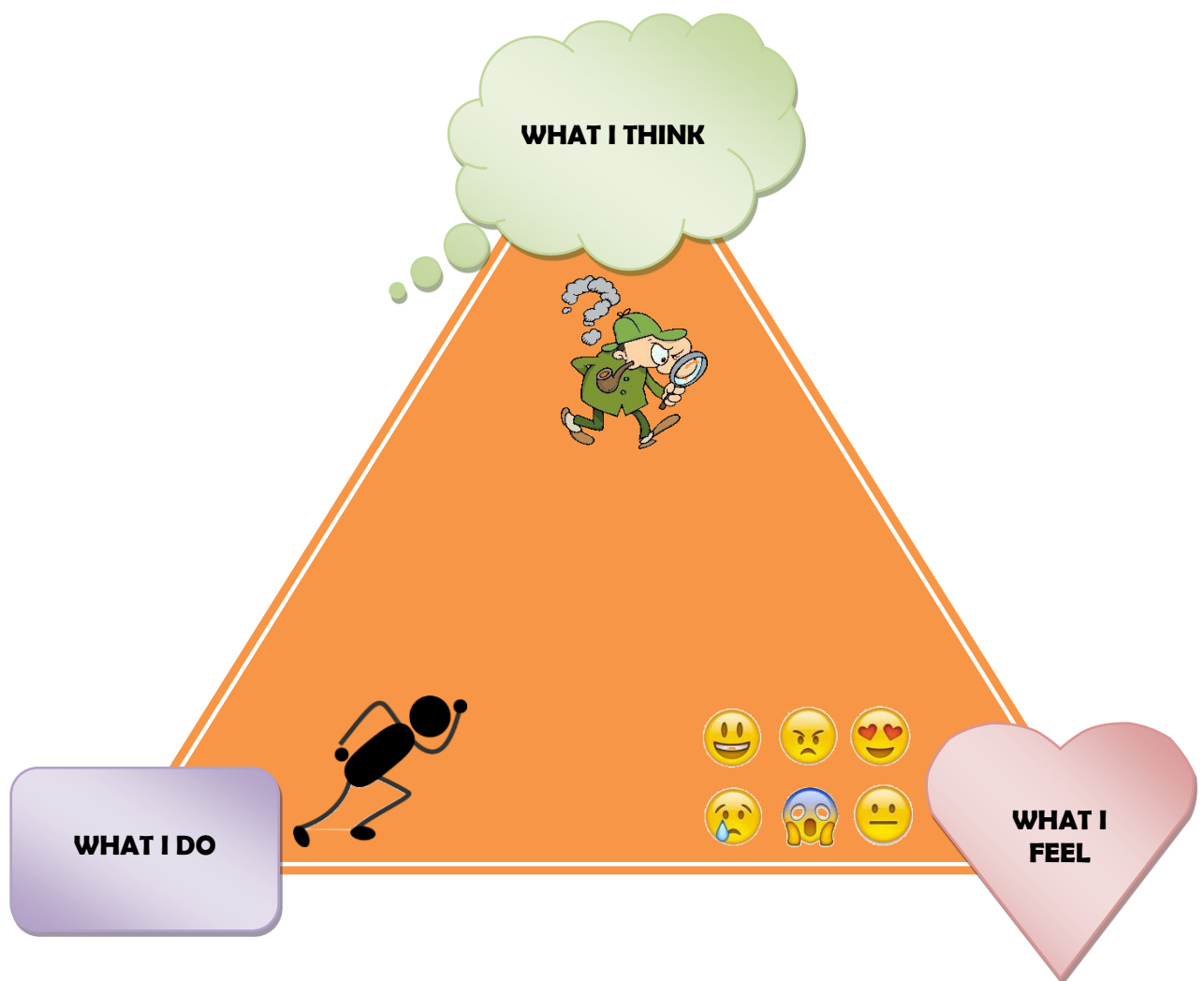
- Duncan, B. L., Sparks, J. A., Miller, S. D., Bohanske, R. T., & Claud, D. A. (2006). Giving youth a voice: A preliminary study of the reliability and validity of a brief outcome measure for children, adolescents and caretakers. *Journal of Brief Therapy, 5*, 71-88.
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*, 169-200.
- Greenberger, D. & Padesky, C. A. (1995). *Mind over mood: A cognitive therapy treatment manual for clients*. New York, NY, US: Guilford Press.
- Kitchens, S. A., Rosén, L. A., & Braaten, E. B. (1999). Differences in anger, aggression, depression, and anxiety between ADHD and non-ADHD children. *Journal of Attention Disorders, 3*, 77-83.
- Novaco, R. W. (2010). Anger and psychopathology. In M. Potegal, G. Stemmler, & C. Spielberger (Eds.). *Handbook of anger* (pp. 465-497). New York: Springer.
- Stallard, P. (2002). *Think Good - Feel Good: A Cognitive Behaviour Therapy Workbook for Children and Young People*. West Sussex, England: John Wiley & Sons Ltd.
- Sukhodolsky, D. G., Kassinove, H., & Gorman. B. S. (2004). Cognitive-behavioral therapy for anger in children and adolescents: A meta-analysis. *Aggression and Violent Behavior, 9*, 247-269.
- Uher, R., Heyman, I., Turner, C. M., & Shafran, R. (2008). Self-, parent-report and interview measures of obsessive-compulsive disorder in children and adolescents. *Journal of Anxiety Disorders, 22*, 979-990.
- Whiteside, S. P., & Abramowitz, J. S. (2005). The expression of anger and its relationship to symptoms and cognitions in obsessive-compulsive disorder. *Depression and Anxiety, 21*, 106-111.

## 8. APPENDICES

---

### 8.1. “Magic Triangle” (Stallard, 2002)

# The Magic Triangle



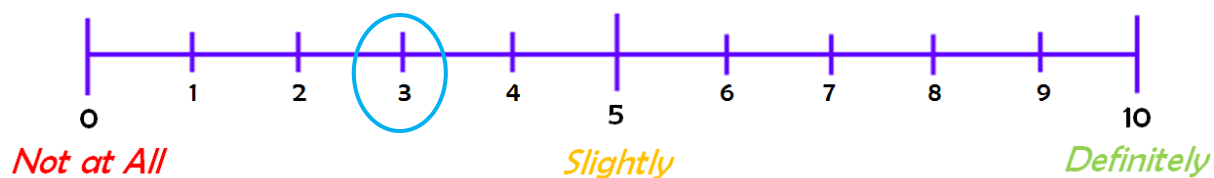
## 8.2. Goal-Setting Worksheet (used in Session 1)



### **Goal 1:**

*"To go from feeling angry 45% of the time to 10%"*

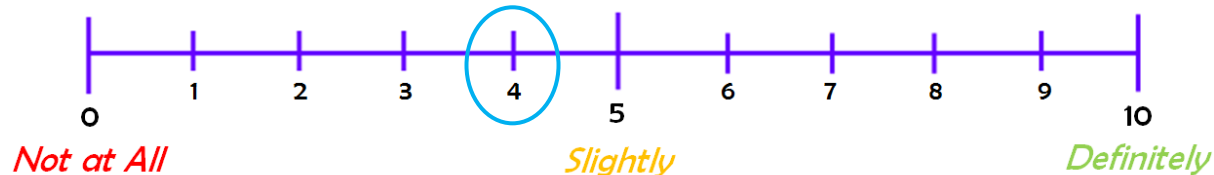
On the scale of 0 – 10, how much do I believe I can do this NOW?



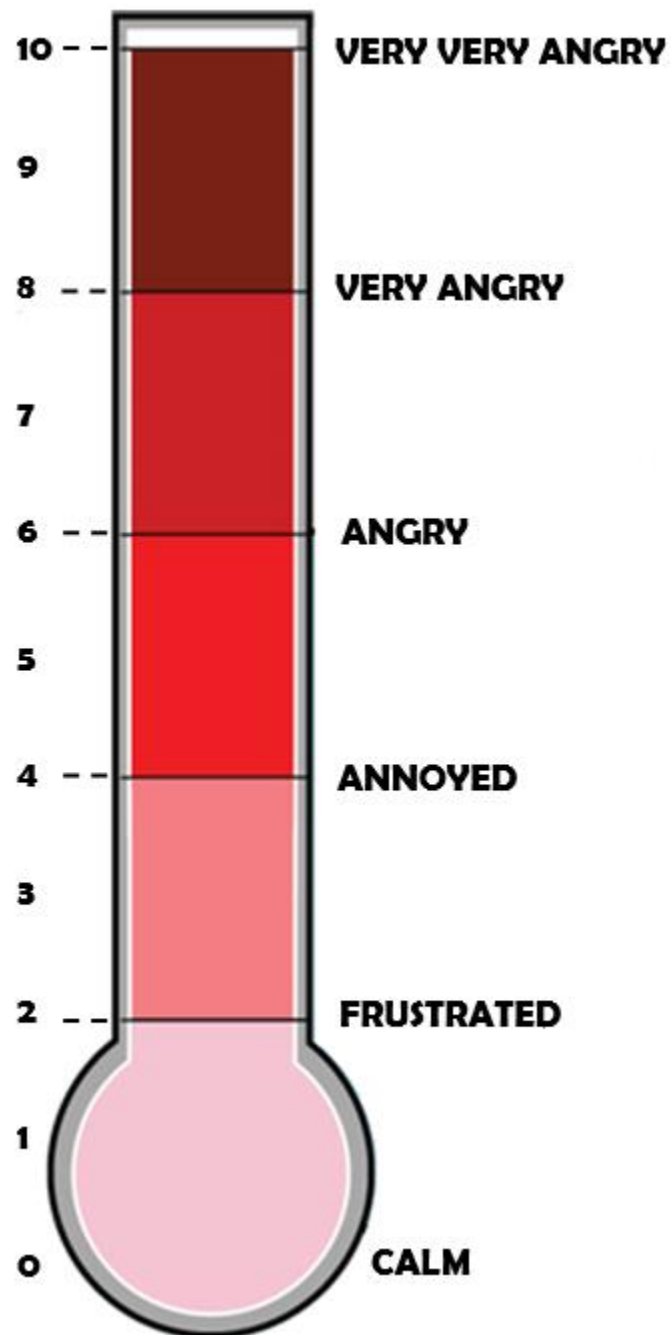
### **Goal 2:**

*"To go from one unhelpful way of coping (shouting at mum) to learning five new helpful ways of coping"*

On the scale of 0 – 10, how much do I believe I can do this NOW?



### 8.3. Anger Thermometer with Tommy's labels (used in session 3)





## 8.4. Reviewed Goals (reviewed in Session 8)



### **Goal 1:**

*"To go from feeling angry 45% of the time to 10%"*

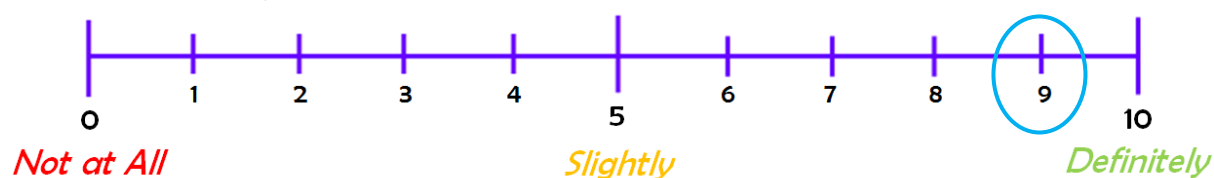
On the scale of 0 – 10, how much do I believe I can do this NOW?



### **Goal 2:**

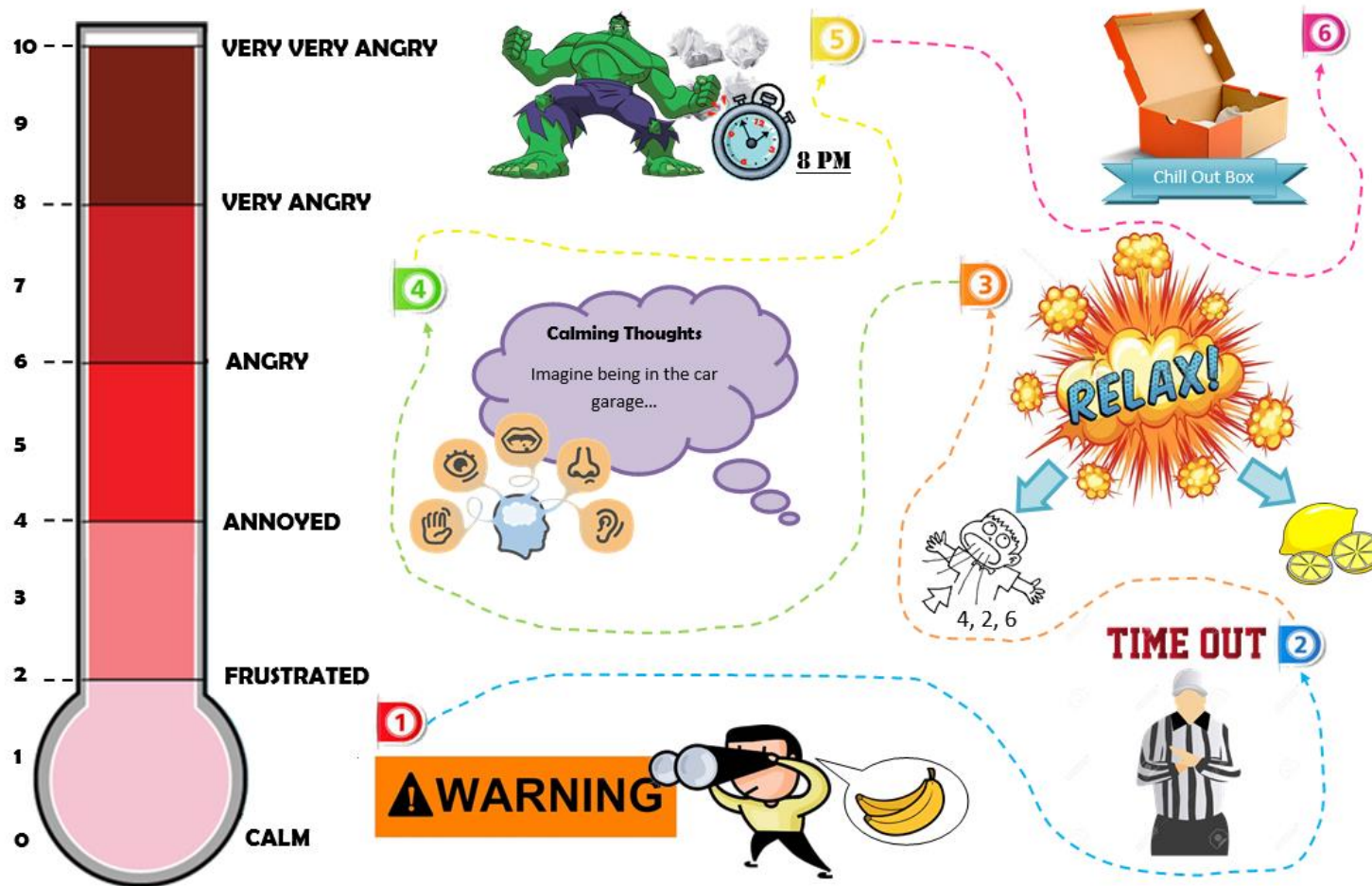
*"To go from one unhelpful way of coping (shouting at mum) to learning five new helpful ways of coping"*

On the scale of 0 – 10, how much do I believe I can do this NOW?



8.5. Relapse Prevention Plan (used in Session 8)

## MY Anger Thermometer & Plan



## CASE STUDY THREE

---

### **Detecting the Dysfunction:**

A neuropsychological assessment of memory with a 74-year-old man diagnosed with Parkinson's disease

---

**Supervised by:** Dr David Matthews

## ABSTRACT

---

“Wilfred” was a 74-year-old gentleman, previously diagnosed with Parkinson’s disease. He was referred to a local Memory Service after reporting memory concerns to his GP. Following an initial memory assessment, there was no evidence of dementia (as indicated by cognitive screening tools). However, due to a high-level of premorbid functioning – demonstrated by his educational and occupational history – Wilfred was referred for neuropsychological testing to obtain a reliable indication of his pre-morbid and current functioning, and to provide more evidence of cognitive decline. Wilfred was seen over three sessions and the neuropsychological assessment battery consisted of tests measuring: premorbid abilities; intellectual functioning; memory abilities; frontal/executive functioning; language; and visual perception. The findings were interpreted as reflecting a focal impairment of a vascular cause, only impacting on Wilfred’s memory for visual stimuli. The deterioration in his visual memory ability was greater than the level expected for a mild cognitive impairment in this domain, but the overall findings were not consistent with a Parkinson’s dementia profile. The limitations and reflections from this case are also discussed.

## CONTENTS

---

<b>1. LITERATURE REVIEW.....</b>	<b>63</b>
1.1. Neuropsychological Assessment in Older Adults.....	63
1.2. Assessing Dementia.....	63
1.3. Differential Diagnosis .....	64
<b>2. CASE DESCRIPTION .....</b>	<b>66</b>
2.1. Background.....	66
2.2. Referral.....	66
2.3. Initial Memory Assessment.....	66
2.4. Cognitive Screening Measures .....	67
<b>3. HYPOTHESES .....</b>	<b>69</b>
<b>4. NEUROPSYCHOLOGICAL ASSESSMENT.....</b>	<b>70</b>
4.1. Behavioural Observations .....	70
4.2. Neuropsychological Assessment Battery .....	70
<b>5. RESULTS.....</b>	<b>73</b>
5.1. Estimated Premorbid Ability .....	73
5.2. Current Intellectual Functioning .....	73
5.3. Memory Ability.....	74
5.4. Frontal/Executive Functioning .....	74
5.5. Language .....	75
5.6. Visual Perception.....	75
<b>6. DISCUSSION .....</b>	<b>76</b>
6.1. Interpretation Summary.....	76
6.2. Limitations.....	76
6.3. Reflections.....	77
<b>7. REFERENCES.....</b>	<b>79</b>
<b>8. APPENDICES .....</b>	<b>82</b>
8.1. Wilfred's Cognitive Profile.....	82
8.2. Conversion of z-scores to percentiles. ....	83

## 1. LITERATURE REVIEW

---

### 1.1. Neuropsychological Assessment in Older Adults

Currently in the National Health Service (NHS), older adults can undergo neuropsychological assessment via three care pathways. That is, (1) clinical neuroscience centres, where referrals are made by a neurologist/neurosurgeon for very specialist care; (2) stroke services, which provide early intervention or rehabilitation to older adults with acute medical conditions; and (3) dementia/memory clinics, which diagnose and manage memory problems within a multidisciplinary team (Morris & Brookes, 2013).

As part of the Department of Health's national strategy to improve the provision for people diagnosed with dementia, this pathway was more defined (Department of Health, 2009). Prior to this, older adult services were typically designed for those at the 'severe' end of the spectrum rather than early intervention. Similarly, the research demonstrated that only a third of people with dementia received a formal diagnosis, which was often given too late (e.g. in crisis) for the individual to have capacity to make appropriate life decisions themselves. Therefore, this strategy aimed to improve national awareness of dementia; provide early diagnosis and intervention; and offer higher quality of care. Subsequently, the number of locally commissioned Memory Services increased, which are now committed to delivering rapid and competent specialist assessment, diagnosis, treatment, and care and support following dementia diagnoses.

### 1.2. Assessing Dementia

Dementia can be considered an umbrella term for a wide range of neurological conditions characterised by a decline in cognitive functioning, with memory deficits being the primary, and arguably the most debilitating, diagnostic feature (Gracey & Morris, 2007). Detecting the early onset of dementia, which by nature is a slow and progressive disorder, is a difficult task, but critical due to a lack of reliable biological markers which can distinguish dementia from normal aging or other neurodegenerative disorders (Salmon & Bondi, 2009).

Neuropsychological assessment is the use of standardised tests to assess cognitive abilities and is one of the tools used within Memory Services to detect dementia (Teng & Manly, 2005). In the absence of explicit tests for dementia, the function of neuropsychological tests are to detect or rule out cognitive impairment, and evaluate whether such performance patterns are 'abnormal' by comparing them to performance profiles empirically seen in other clinical

disorders. These cognitive patterns are also key in the management of the disorder; however, there is a danger of false-positives, which is why neuropsychological tests are not sufficient alone in concluding diagnoses (Morris & Brookes, 2013). Jacovaa and colleagues (Jacovaa, Kertesz<sup>b</sup>, Blair<sup>b</sup>, Fisk<sup>c</sup> & Feldman<sup>a</sup>, 2007) suggest that neuropsychological testing should be applied selectively, and only to address: distinctions between age-appropriate cognitive decline, Mild Cognitive Impairments (MCI), or dementia; the risk of someone with a MCI progressing to a dementia; and the presence of any differential diagnoses of cognitive impairment. Particularly in the early stages of dementia however, subtle complaints in functioning are important to investigate as they may have significant yet unrecognised impact on functioning (Sano, 2006).

To meet diagnostic criteria for dementia, an individual is required to demonstrate impairments across several cognitive domains, so not only is this for memory, but also in one or more of the following areas: language, praxis, gnosis, and executive function (Teng & Manly, 2005). Consequently, neuropsychological testing batteries are required to incorporate a combination of these measures, and not exclusively assess memory. Nonetheless, if the cognitive dysfunction observed is limited to only one domain and does not significantly interfere with functioning, this does not warrant a dementia diagnosis, and MCI may be more appropriate. It may also be necessary to repeat neuropsychological testing for the more diagnostically difficult cases, as this can assess one of the key features of dementia - cognitive decline over time (Gerrard, 2008).

### 1.3. Differential Diagnosis

Memory problems can be explained by several causes/diagnoses, therefore part of the assessment process is to consider the most appropriate explanation, as this will directly influence treatment and management. Firstly, there are many variants of dementia (e.g. Alzheimer's, vascular, frontal-temporal, Parkinson's), and so clinicians need to be aware of their differing presentations, both generally and specifically within neuropsychological testing. For instance, on a test of verbal fluency, an individual with Alzheimer's disease would typically spend longer retrieving words beginning with a specified letter compared to words within a category and would frequently repeat their responses. An individual with frontal-temporal dementia however, would involuntarily repeat words and be unable to inhibit a tendency to provide obscenities or words with sexual connotations (Gerrard, 2008).

Secondly, a key consideration is whether the cognitive dysfunction observed is secondary to a neurodegenerative process, or a symptom of a more generalised underlying condition which can subsequently be treated (Gerrard, 2008). Some of the main differential diagnoses are:

depression, delirium, and use of psychoactive substances (including alcohol) (Bottino et al., 2011). Depression is often mistaken for dementia, and so this should be clinically assessed in all clients, although Mully (1986) and Bottino et al. (2011) have suggested some observational differences to help determine the influence of depression on memory. That is, compared to Alzheimer's disease, an individual with depression is likely to complain about the memory difficulties; perform better than they self-report; perform better when cues are offered; respond with "I don't know" to direct questioning; reduce their effort with cognitive demand; communicate their distress; and will not attempt to conceal the difficulties.

Alternatively, there are other pertinent factors in the older adult population that can influence memory abilities, including poor physical health and negative side-effects from subsequent medications. Physical disabilities may place the individual at a disadvantage on neuropsychological tests that require motor abilities, especially if they are time-limited. Adaptions may also be required for those with visual/hearing impairments. Medications, such as benzodiazepines, antiepileptic and anti-parkinsonian drugs, and anti-hypertensive agents, can also cause intellectual impairment, although positively, these effects can be reversed when the medication is withdrawn (Bottino et al., 2011; Mulley, 1986). Blood tests can be a good method of eliminating reversible physical explanations associated with cognitive dysfunction, for example anaemia, uraemia, and vitamin B deficiencies (Gerrard, 2009). Further, this population may be more exposed to stressors that impact upon their mental health, including the loss of friends/family and poor physical health, or as with any age group, they may engage in a lifestyle predictive of memory dysfunction (e.g. alcohol/drug abuse, poor diet and exercise, smoking).

Taken together, diagnosing dementia is a complicated process of exclusion, whereby it is required that all other possible explanations for the cause of cognitive dysfunction (including alternative sub-types of dementia, MCI, psychiatric disorders, physical health, and lifestyle) are ruled out before a diagnosis is given. This reinforces the importance of multidisciplinary investigations to indicate which prognosis is more/less probable, such as head scans, blood tests, observation, and clinical interviews, whereby neuropsychological assessment is just one piece of this diagnostic puzzle.



## 2. CASE DESCRIPTION

---

### 2.1. Background

'Wilfred' (pseudonym) was a 74-year-old, white British male, who previously worked as a Violinist and Musical Conductor/Arranger. He had since retired, although still leisurely engaged in various musical pursuits. He was also married and had two sons.

### 2.2. Referral

Wilfred was referred to his local Memory Service in January 2016 by his GP, who Wilfred had reported memory concerns to. Wilfred had noticed that over the previous two/three years, he was more frequently forgetting things and was only able to remember them when he wrote them down.

### 2.3. Initial Memory Assessment

Following his referral, Wilfred was assessed by a member of the multidisciplinary team. The purpose was to obtain a first-hand account of Wilfred's subjective memory concerns, and to evaluate the extent and interference of these difficulties. Moreover, this meeting allowed for a brief risk assessment to be conducted, as well as collecting relevant background information.

Wilfred reported mild memory problems, such as misplacing items at home and forgetting recent events/people he had met. He described often forgetting to take his medications and leaving tasks unfinished. He also noticed that it was becoming harder for him to fill in paperwork/forms and read numbers, and experienced some word finding difficulties. However overall, these problems did not appear to be significantly impacting on his day-to-day functioning. No collateral account could be obtained, although Wilfred did report that his wife and son had confirmed that they have also noticed these changes.

Wilfred grew up and was educated locally, passing both his O and A-Levels, subsequently progressing to obtain a 4-year music degree at the University of Cambridge. Following this, he studied violin for another two years and worked as a violinist and conductor/musical arranger. He reported no known family history of memory problems. In 2002, Wilfred was diagnosed with Parkinson's disease. Prior to this (late 1990's), Wilfred reported one episode of depression, which was treated successfully with antidepressants. He also described experiencing a one-off panic attack approximately two years ago but attributed this to a sudden onset of unusual body

sensations following a medication change. He currently described a stable mood with no concerns regarding his mental health. Wilfred was a non-smoker, although would drink alcohol occasionally (1-2 drinks per evening) and reported historical cocaine use in the early 1980's. Wilfred was still independent of all activities of daily living, and remained socially active, participating in several hobbies. In terms of risk, Wilfred reported no suicidal thoughts; no incidents of fire/floods around the home, nor wandering or getting lost; he no longer drove; and had no access to children or vulnerable adults. He was considered to be a low risk of falls due to his Parkinson's diagnosis and a reported "trip" approximately 12-18 months previously.

## 2.4. Cognitive Screening Measures

On assessment, the service requires the administration of routine screening tools to identify the likelihood of clinically significant cognitive problems. They provide a standardised measure of the memory difficulties and assess other differential diagnoses (anxiety/depression). These screening measures have been outlined in Table 5 below, along with Wilfred's scores.

**Table 5: A list and description of the Cognitive Screening Measures used, along with Wilfred's score.**

Screening Measure	Description	Wilfred's Score
<b>The Standardised Mini-Mental State Examination (SMMSE)</b> <i>(Molloy, Alemayehu, &amp; Roberts, 1991)</i>	The SMMSE screens for cognitive impairment in the elderly across a number of domains (orientation to time and place, immediate recall, short-term memory, calculation, language, and constructive ability). The client can obtain a maximum score of 30, and scores of 24 or less are suggestive of dementia.	28 / 30  Lost points on the 'figure copy' and 'phrase repetition'.
<b>The Addenbrooke's Cognitive Exam – 3<sup>rd</sup> Version (ACE-III)</b> <i>(Adaption of ACE-Revised (ACE-R), Mioshi, Dawson, Mitchell, Arnold, &amp; Hodges, 2006)</i>	The ACE-III is a brief cognitive screening tool used for the detection and differentiation of dementias. This tool is scored out of 100, with higher scores indicating better functioning. Five domains contribute to this total score, which are: attention and orientation, memory, verbal fluency, language, and visuo-spatial skills. Scores of 88 and 82 out of 100 are used as a cut-off to indicate the likelihood of dementia.	94 / 100  Lost one point on the 'attention and orientation' domain, and five on the 'visuospatial' domain.

Screening Measure	Description	Wilfred's Score
<b>The Hospital Anxiety and Depression Scale (HADS)</b> <i>(Zigmond &amp; Snaith, 1983)</i>	The HADS was initially developed to screen for anxiety and depression among clients in non-psychiatric settings, however it is now used in various clinical settings. It is a self-report measure consisting of 14 questions, where the client is required to rate the presence of each symptom on a 4-point likert scale. From this, an anxiety and depression subscale score can be derived (both maximum scores = 21), with scores of 11 and over indicating clinical significance.	Anxiety = 4 / 21  Depression = 3 / 21

### 3. HYPOTHESES

---

From the assessment, several hypotheses could be considered regarding the cause of Wilfred's subjective memory concerns. Some of the differential diagnoses include Wilfred's lifestyle (alcohol intake) and history of depression. However, due to the increased risk of developing dementia in individuals with Parkinson's disease, this was one of the main working hypotheses. Moreover, the experience of Parkinson-related symptoms and taking the medication alone could be cause of disturbances in his processing. Alternatively, the memory difficulties might best be explained by a MCI or appropriate age-related decline.

From the cognitive screening tools however, Wilfred's scores were all within a 'healthy' range, suggesting that a diagnosis of dementia was not likely, and that his mood was not having a negative effect on his memory. Nevertheless, it should be noted that these measures have been found to not be sensitive to detecting cognitive impairments in higher functioning individuals, and therefore Wilfred's previous high level of functioning (as indicated by his high academic achievements) may have skewed these results, not allowing for an accurate reflection of the subjective decline in memory that Wilfred was describing.

Due to these reasons, Wilfred was placed on the waiting list for a neuropsychological assessment to obtain a reliable indication of his pre-morbid and current functioning, and to provide more evidence of cognitive decline.

## 4. NEUROPSYCHOLOGICAL ASSESSMENT

---

### 4.1. Behavioural Observations

Wilfred attended three sessions at his local Memory Service. The assessment consisted of two standard sessions, where the service's standard battery of tests was administered. After reviewing the initial results, an additional session was scheduled to obtain more specific information regarding particular cognitive domains using additional tests.

Wilfred was assessed in a quiet room, free from distractions, and he appropriately used his reading glasses as needed. He did not appear to show signs of anxiety and reported being happy to complete all the tasks. Wilfred demonstrated that he was giving thought to his answers, as he would correct himself, ask appropriate questions, and even provide responses to earlier questions once he had thought of the intended answer. He appeared able to remain focused for the full duration of each task, and demonstrated a determination to perform well, often requesting feedback on his performance. The conversation was easy to follow, although at times clarity was needed as Wilfred's speech was soft and the pronunciation was not clear. Wilfred's diagnosis of Parkinson's disease could also affect his leg movements, but there were no concerns with his fine motor skills, as evident through his ability to appropriately manipulate the blocks in the 'Block Design' sub-test. He was also orientated to person/time/place.

### 4.2. Neuropsychological Assessment Battery

Table 6 below provides a description of each test that was used in Wilfred's cognitive assessment, as well as indicating whether it formed part of the standard testing battery, or if it was used as an additional test.

**Table 6: A list and description of the neuropsychological tests used (Strauss, Sherman & Spreen, 2006).**

Cognitive Domain	Test used	Test Description
<b>Standard Battery</b>		
<b>Intellectual Functioning</b>	- Test of Premorbid Functioning (TOPF) <i>(Wechsler, 2011<sub>a</sub>)</i>	- The participant is required to pronounce a list of 70 words that have atypical grapheme to phoneme translations, an ability that has been shown to remain stable even in the presence of cognitive decline. The overall score determines their estimated level of premorbid functioning.
	- Wechsler Abbreviated Scale of Intelligence 2 <sup>nd</sup> Version (WASI-II) <i>(Wechsler, 2011<sub>b</sub>)</i>	- This provides a reliable and brief estimate of current intellectual functioning. It consists of 4 sub-tests, half of which assess verbal abilities, and the other half non-verbal.
<b>Parietal Functioning</b>	- Wechsler Adult Intelligence Scale – 4 <sup>th</sup> Version (WAIS-IV) ‘arithmetic’ sub-test <i>(Wechsler, 2008)</i>	- The participant is required to mentally solve arithmetic word problems that are presented orally within a time-limit.
<b>Memory</b>	- California Verbal Learning Test (CVLT) <i>(Delis, Kramer, Kaplan, &amp; Ober, 2000)</i>	- The CVLT measures verbal learning and memory by asking the participant to immediately recall a list of words over multiple trials. Participants are also asked to recall these words following a distractor task and after a delay of 20 minutes.
	- Rey Osterreith Complex Figure Test <i>(Meyers &amp; Meyers, 1995)</i>	- This test requires the participant to copy a complex geometric design, immediately recall it, and recall it again following a 30-minute delay. It measures visual-spatial construction ability and visual memory.
<b>Executive Functioning</b>	- Letter and Category Verbal Fluency (FAS) <i>(Gladsjo, Schuman, Miller &amp; Heaton, 1999)</i>	- The FAS measures a participant’s spontaneous production of words under two restricted search conditions. That is (1) letters (phonemic), and (2) category (semantic). Participants are required to verbally provide as many words associated with that condition as possible, within a specified time limit (one-minute).
	- Hayling Sentence Completion Test <i>(Burgess &amp; Shallice, 1997)</i>	- This test measures behavioural regulation, initiation speed and response suppression. It consists of two sets of 15 sentences where the last word is missing. Participants are firstly asked to provide a meaningful word that completes the sentence, and in the second part, participants are asked to provide a word that is unconnected to the sentence in every way. Participants are also required to provide their answers as quickly as possible for both parts of the test.

Cognitive Domain	Test used	Test Description
	- Trail Making Test ( <i>Reitan, 1958</i> )	- This test assesses attention, speed, and mental flexibility, and is split into two parts (A & B). Part A requires the participant to draw one continuous line connecting 25 randomly positioned numbers together in numerical order. Part B requires the participant to repeat this task, but alternate between numbers and letters.
<b>Language</b>	- Graded Naming Test ( <i>McKenna &amp; Warrington, 1983</i> )	- This test assesses object-naming ability and can detect word-finding difficulties. Participants are required to name a variety of object drawings that are also graded in difficulty.
<b>Additional Battery</b>		
<b>Memory</b>	- Wechsler Memory Scale (WMS-IV) ( <i>Wechsler, 2009</i> )	- The WMS-IV is a comprehensive battery of tests that assesses verbal and non-verbal memory abilities in adults across four indices: auditory, visual, immediate, and delayed memory.
<b>Visuo-spatial Functioning</b>	- Visual Object and Space Perception Battery (VOSP) ( <i>Warrington &amp; James, 1991</i> )	- This test consists of eight separate tasks which are designed to measure object and spatial perception, while minimising the involvement of other cognitive skills. Four tasks are dedicated to each domain.

## 5. RESULTS

---

Please see Appendix 8.1 for Wilfred's cognitive profile and raw scores.

When reporting Wilfred's scores, a "percentile" rank is given to demonstrate Wilfred's rank in the national comparison group. If the percentile rank was in the 45<sup>th</sup> percentile for example, it would indicate that Wilfred scored higher than approximately 45 out of 100 adults his age. Percentiles are a common measure used by many standardised neuropsychological tests; therefore, they are part of the test's interpretation process and can be calculated directly from the test's raw scores. However, with some tests, the direct conversion into a percentile is not provided. In these cases, scores are converted into "z-scores" (aka. "standard scores") which represent how much an individual's score deviates from the norm. Z-scores can be calculated using the participant's score and the mean and standard deviation from the comparison group. These z-scores can then be directly converted to percentiles using conversion tables. For example, a z-score of 0 would represent no deviation from the group mean, and therefore would convert to a percentile of 50 (i.e. the average). A positive z-score represents performance above the norm, and negative scores represent performance below the norm. Please see Appendix 8.2 for an illustration of how z-scores convert to percentiles.

### 5.1. Estimated Premorbid Ability

Wilfred was able to read 66 out of 70 words correctly on the TOPF. This provided an estimated premorbid verbal IQ of 123 (94<sup>th</sup> percentile), a performance IQ of 122 (93<sup>rd</sup> percentile) and a full-scale IQ score of 125 points, placing him at the 96<sup>th</sup> percentile. These scores can be considered in the "superior" range and are consistent with his educational and occupational attainment.

### 5.2. Current Intellectual Functioning

Employing the WASI-II, Wilfred obtained a verbal IQ of 105 points (63<sup>rd</sup> percentile) and a performance IQ of 100 points (50<sup>th</sup> percentile), creating a full-scale IQ of 103 points (58<sup>th</sup> percentile). These scores are less than those predicted by TOPF, and overall are within the "average" range.

On the WAIS-IV 'arithmetic' sub-test, Wilfred achieved a scaled score of 16 out of 19, placing his abilities at the 98<sup>th</sup> percentile and in the "superior" range.



### 5.3. Memory Ability

#### **Verbal Memory:**

With regards to the CVLT, on the parameter of the single list recall Wilfred successfully recalled 9 of a possible 16 words presented. In relation to age and sex norms, such a score places his abilities at the 99.9<sup>th</sup> percentile. Across five presentations of the material he was able to generate a total of 60 out of a possible 80 words. Such a score places his current abilities at the 99<sup>th</sup> percentile. Following a distracter task, he was able to recall 14 of a possible 16 words presented, such a score placing his abilities at the 99<sup>th</sup> percentile. Following a delay of approximately 20 minutes, Wilfred was able to recall 14 of a possible 16 words placing his abilities at the 98<sup>th</sup> percentile.

#### **Visual Memory:**

On the Rey Osterreith Complex Figure Test, Wilfred proved able to successfully copy 28.5 of a possible 36 items presented, such a score placing his abilities in relation to appropriate norms at the 24<sup>th</sup> percentile. Wilfred was able to immediately recall 6.5 items of a possible 36, placing his at the 7<sup>th</sup> percentile. Following a delay of approximately 30 minutes, Wilfred was able to recall 10 of the 36 items previously presented, placing his abilities at the 21<sup>st</sup> percentile.

#### **General Memory Abilities:**

On the WMS-IV test, Wilfred's auditory memory performance was scored 130 (98<sup>th</sup> percentile) and his visual memory performance was scored 87 (19<sup>th</sup> percentile). His immediate memory abilities were scored at 110 (75<sup>th</sup> percentile), and his delayed memory abilities were 119 (90<sup>th</sup> percentile).

### 5.4. Frontal/Executive Functioning

When the FAS test was employed, Wilfred generated a total of 66 words across the three categories of letters specified, such a score placing his abilities at the 98<sup>th</sup> percentile. When animals were employed as a semantic category, Wilfred was able to successfully generate 20 exemplars, placing his abilities at the 66<sup>th</sup> percentile.

On the Hayling subtest, Wilfred obtained an overall scaled score of 7 out of 10, indicating his performance was in the "high average" range. Upon part two of the test, Wilfred obtained a scaled score of 7 out of 8, indicating his inhibitory control was also in the "high average" range.

Upon Trail A of the Trail Making Test, Wilfred completed the paradigm in 30 seconds with no errors, placing his abilities at the 79<sup>th</sup> percentile. He completed Part B of the test in 89 seconds with one error, placing his abilities at the 62<sup>nd</sup> percentile.

### 5.5. Language

The Graded Naming Test was employed as a measure of confrontational naming. Wilfred successfully named 24 of 30 items, placing him at the 82<sup>nd</sup> percentile.

### 5.6. Visual Perception

Wilfred passed seven out of eight subtests from the VOSP, where he did not reach the pass criteria for the 'Progressive silhouettes' test. However due to an administration error during this test, the testing material was not rotated as required which may have had a negative impact on his results and therefore placed Wilfred at a disadvantage.

## 6. DISCUSSION

---

### 6.1. Interpretation Summary

The neuropsychological assessment revealed that Wilfred's premorbid functioning was in the "superior" range, whereas his current intellectual abilities were in the "average" range, somewhat lower than what would be expected. Therefore, there is evidence of decline in his overall intellectual ability. The findings also consistently showed no concerns regarding Wilfred's memory of verbal stimuli, as evident through his "superior" performance on these sub-tests. Yet in comparison, his memory ability for visual stimuli was significantly weaker, as demonstrated by his "borderline-low average" performance. On further exploration, the testing confirmed no difficulties with Wilfred's perception of visual objects or space, and despite poor immediate and delayed memory for visual items, his recognition of these items was good and intact. Wilfred also demonstrated adequate language and executive functioning abilities.

These results appear to reflect a focal impairment, only impacting on Wilfred's memory for visual stimuli. The deterioration in his visual memory ability was greater than the level expected for a MCI in this domain, but the overall findings were not consistent with a Parkinson's dementia profile. It is also possible that the findings represented a congenital weakness that Wilfred always had, so may not necessarily represent an organic problem/cause. Together, the findings best fitted with a vascular cause; although it was recommended that more evidence was required before for a formal diagnosis was made.

### 6.2. Limitations

It can be argued that the ecological validity of the tests administered is limited, and consequently did not provide a true representation of the extent of memory difficulties experienced. For instance, individuals are not usually required to learn comprehensive lists of words day-to-day or have to remember complex/unusual figures. Similarly, although the tests used in this assessment are standardised, the quality of this process is variable from test-to-test. For example, the interpretations of findings are based on comparisons to the available normative data. However, these norms are typically based on small non-British samples, with little discrimination between demographic variables. Therefore, the generalisability to an individual such as Wilfred is limited.

One domain that this neuropsychological assessment did not measure was Wilfred's effort and motivation to perform to his best ability. The British Psychological Society (2009) outlines how clinical judgement alone is a poor estimate of effort and suggests that best practice is to incorporate a standalone or embedded effort test. As this was not completed with Wilfred, it raises questions about the validity of the interpretations drawn, and ultimately the recommended diagnosis. Contrastingly, Rudman, Oyebode, Jones, and Bentham (2011) found that performance on effort tests were related to cognitive functioning in the working age dementia population, whereby those with moderate/severe dementia performed worse on effort tests compared to a mild group, and the effort tests themselves were equally not sensitive to cognitive dysfunction. Anecdotally, clinicians have suggested that the inclusion of effort tests can increase patient burden and fatigue, as well as require more of the limited time and resources that are available (in non-medico legal/forensic contexts). Nies and Sweet (1994) further discourage the use of one test to measure effort, as this only draws conclusions about effort in one cognitive domain. Therefore, they suggest that multiple tests should be adopted to clearly distinguish between effort and cognitive factors, which would further add to this burden and demand on clinical resources. Subsequently, when a choice is offered, effort tests were usually not administered in this service, however since completing this case, these are being made compulsory.

### 6.3. Reflections

I enjoyed the "detective" nature of neuropsychological assessment and got pleasure out of finally "solving the problem" for the client. However, I found that my interpretations were easily influenced by my initial hypothesis. Supervision helped my awareness of this, highlighting the importance of taking a neutral/unbiased stance, and how the role of neuropsychological assessments is to rule out disorders and not to rule them in. Additionally, neuropsychological tests are a valuable source of information and hold significant weight in the assessment process. Therefore, I chose not to view the results of Wilfred's MRI scan until after the report had been finalised, as I did not want to contaminate my original interpretation.

This scan ultimately came back as "normal" and the multidisciplinary team concluded that the results were indicative of a MCI in the context of Parkinson's disease. Therefore, it was likely Wilfred's memory difficulties would fluctuate in line with symptoms and medication. I was initially surprised to learn the results of Wilfred's scan, especially as it felt that his memory for visual information was so profound. This made me consider the limitations of brain scans, and how there was no premorbid measure available, where I would have estimated that like the

TOPF, this would have also been significantly above the norm. Subsequently, there could have been deterioration; however, this was not picked up due to the current measure falling within the “normal” range. This emphasised the added proficiency of neuropsychological assessment, and I learnt how it is common to experience inconsistencies within the multidisciplinary evidence.

I also found it difficult not being involved in the feedback process of the neuropsychological assessment results, as it made me feel as though I had not met my client’s needs - particularly as I identified with Wilfred’s competitive side and his desire to know how well he performed on each test. This responsibility is given to the care coordinator, where the feedback would not have been specific to the neuropsychological assessment, but rather the overall conclusion from the multidisciplinary team. I can’t help but be dissatisfied with this service arrangement considering the huge efforts that Wilfred put in when undergoing the testing. I would have liked the opportunity to summarise the results and their meaning to Wilfred myself, especially as he was interested in this. Moreover, I would have been able to draw on his personal strengths/weaknesses and subsequently make personalised recommendations for how he could manage the subjective difficulties day-to-day and enhance memory.

## 7. REFERENCES

---

- Bottino, C. M. C., de Pádua, A. C., Smid, J., Areza-Fegyveres, R., Novaretti, T., Bahia, V. S., & Working Group on Alzheimer's Disease and Vascular dementia of the Brazilian Academy of Neurology. (2011). Differential Diagnosis between Dementia and Psychiatric Disorders. *Dementia e Neuropsychologia*, 5, pp. 288-296.
- Burgess, P. & Shallice, T. (1997) *The Hayling and Brixton Tests: Test manual*. Bury St Edmunds: Thames Valley Test Company.
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test – Second Edition*. San Antonio, TX: Psychological Corporation.
- Department of Health. (2009). *Living Well with Dementia: A National Dementia Strategy*. Retrieved from [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/168220/dh\\_094051.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/168220/dh_094051.pdf).
- Gerrard, P. (2008). Differential Diagnosis in Dementia. In S. F. Cappa, J. Abutaleb, j. Démonet, P. C. Fletcher, & P. Gerrard (Eds.), *Cognitive Neurology: A Clinical Textbook* (pp. 185-198). New York: Oxford University Press Inc.
- Gladsjo, J. A., Schuman, C. C., Miller, S. W., & Heaton, R. K. (1999). *Norms for letter and category fluency: Demographic corrections for age, education, and ethnicity*. Odessa, FL: PAR.
- Gracey, D. J. & Morris, R. G. (2007). Neuropsychological assessment in dementia. *Psychiatry*, 6, pp. 498-502.
- Jacovaa, C., Kertesz, A., Blair, M., Fisk, J. D., & Feldman, H. H. (2007). Neuropsychological testing and assessment for dementia. *Alzheimer's & Dementia*, 3, pp. 299 –317.
- McKenna, P. & Warrington, E. K. (1983). *Graded Naming Test*. Windsor: NFER- Nelson.
- Meyers, J. E., & Meyers, K. R. (1995). *Rey complex figure test and recognition trial: Professional manual*. London: PAR, Inc.
- Mioshi, E., Dawson, K., Mitchell, J., Arnold, R., & Hodges, J. R. (2006). The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *International Journal of Geriatric Psychiatry*, 21, pp. 1078-1085.

- Molloy, D. W., Alemayehu, E., & Roberts, R. (1991). Reliability of a standardised Mini-Mental State Examination compared with the traditional Mini-Mental State Examination. *American Journal of Psychiatry*, 148, pp. 102-105.
- Morris, R. G. & Brookes, R. L. (2013). Neuropsychological Assessment of Older Adults. In L. H. Goldstein & J. E. McNeil (Eds.), *Clinical Neuropsychology: A practical Guide to Assessment and Management for Clinicians, Second Edition* (pp. 347-374). West Sussex: John Wiley & Sons LTD.
- Mulley, G. P. (1986). Differential Diagnosis of Dementia. *British Medical Journal*, 292, pp. 1416-1418.
- Nies, K. J., & Sweet, J. J. (1994). Neuropsychological assessment and malingering: A critical review of past and present strategies. *Archives of Clinical Neuropsychology*, 9, pp. 911-922.
- Reitan, R. M. (1958). Validity of the Trail Making test as an indicator of organic brain damage. *Perceptual and Motor Skills*, 8, 271-276.
- Rudman, N., Oyeboade, J. R., Jones, C. A., & Bentham, P. (2011). An investigation into the validity of effort tests in a working age dementia population. *Aging & Mental Health*, 15, pp. 47-57.
- Salmon, D. P. & Bondi, M. W. (2009). Neuropsychological Assessment of Dementia. *Annual Review of Psychology*, 60, pp. 257-282.
- Sano, M. (2006). Neuropsychological Testing in the Diagnosis of Dementia. *Journal of Geriatric Psychiatry and Neurology*, 19, pp. 155-159.
- Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary (3<sup>rd</sup> Edition)*. Oxford: Oxford University Press.
- Teng, E. L. & Manly, J. J. (2005). Neuropsychological testing: Helpful or harmful? *Alzheimer's Disease and Associated Disorders*, 19, pp. 267-271.
- The British Psychological Society. (2009). *Assessment of Effort in Clinical Testing of Cognitive Functioning for Adults*. Retrieved from [http://www.bps.org.uk/sites/default/files/documents/assessment\\_of\\_effort\\_in\\_clinical\\_testing\\_of\\_cognitive\\_functioning\\_for\\_adults.pdf](http://www.bps.org.uk/sites/default/files/documents/assessment_of_effort_in_clinical_testing_of_cognitive_functioning_for_adults.pdf).

- Warrington, E. K. & James, M. (1991). *The Visual Object and Space Perception Battery*. Bury St Edmunds: Thames Valley Test Company.
- Wechsler, D. (1998). *WMS-III UK administration and scoring manual*. London: Psychological Corporation.
- Wechsler, D. (2008). *Manual for the Wechsler Adult Intelligence Scale–Fourth Edition*. San Antonio, TX: Pearson.
- Wechsler, D. (2009). *Wechsler Memory Scale-Fourth Edition*. San Antonio, TX: Pearson.
- Wechsler, D. (2011<sub>a</sub>). *The Test of Premorbid Functioning (TOPF)*. San Antonio TX: The Psychological Corporation.
- Wechsler, D. (2011<sub>b</sub>). *Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI-II)*. San Antonio, TX: NCS Pearson.
- Zigmond, A. S. & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, 67, pp. 361–70.



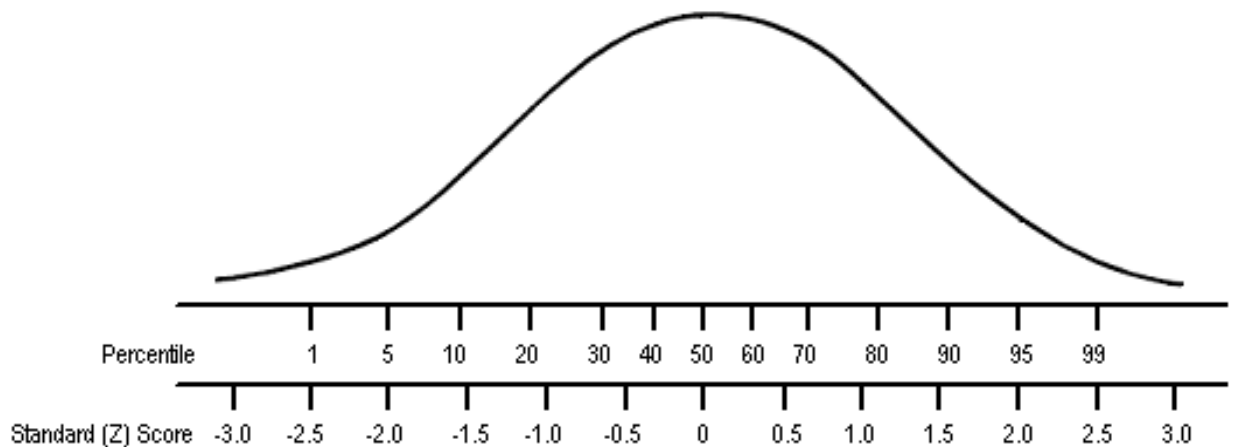
## 8. APPENDICES

### 8.1. Wilfred's Cognitive Profile

DEMENTIA SCREEN:		SMMSE: 28 / 30			ACE III: 94 / 100			
		Scores		Percentile Rank		Classification		
INTELLECTUAL ABILITY IQ								
		VIQ	PIQ	FSIQ	VIQ	PIQ	FSIQ	
PREMORBID – TOPF: (Raw score = 66/70)		123	122	125	94	93	96	Superior (FSIQ)
CURRENT – WASI II:		105	100	103	63	50	58	Average (FSIQ)
WAIS IV: Arithmetic sub-test Scaled score		16 / 19			98			Superior
MEMORY								
Verbal – CVLT:	Single List Recall	9 / 16			99.9			Very Superior
	Learning Curve	60 / 80			99			Superior
	Recall following distraction	14 / 16			99			Superior
	Recall following delay	14 / 16			98			Superior
Visual – Rey:	Copy of Design	28.5 / 36			24			Low Average
	Immediate recall	6.5 / 36			7			Borderline
	Recall following delay	10 / 36			21			Low Average
WMS-IV	Auditory Memory Index	130			98			Superior
	Visual Memory Index	87			19			Low Average
	Immediate Memory Index	110			75			Above Average
	Delayed Memory Index	119			90			Above Average
LANGUAGE								
Graded Naming Test:		24 / 30			82			Above Average
FRONTAL/EXECUTIVE FUNCTIONING								
Controlled Word Association:	FAS	66			98			Superior
	Animals	20			66			Average
Hayling:	Overall Scaled score	7 / 10						High Average
	Section 2 errors Scaled score	7 / 8						High Average
Trails:	A	30 secs (0 errors)			79			Above Average
	B	89 secs (1 error)			62			Average

		Scores	Percentile Rank	Classification
VISUAL OBJECT/SPATIAL PERCEPTION				
<b>VOSP – Object:</b>	Incomplete Letters	19 / 20		Pass
	Silhouettes	22 / 30		Pass
	Object Decision	17 / 20		Pass
	Progressive silhouettes	18 / 20		Fail
<b>VOSP – Spatial:</b>	Dot Counting	10 / 10		Pass
	Position Discrimination	20 / 20		Pass
	Number Location	10 / 10		Pass
	Cube Analysis	9 / 10		Pass

## 8.2. Conversion of z-scores to percentiles.



## CASE STUDY FOUR



### **Widening the Access to “CUES-Ed”:**

Using an adapted a CBT school-based intervention  
for children with Social, Emotional, and Mental  
Health (SEMH) needs

---

**Supervised by:** Dr Anna Redfern

## ABSTRACT

---

“CUES-Ed” is a universal psychoeducation package for primary school children, aimed at improving emotional wellbeing and resilience, reducing stigma, and normalising unusual experiences. This early intervention programme was subsequently adapted and delivered to two classes ( $N = 14$ , one female) in a school for children with Social, Emotional, and Mental Health (SEMH) needs (aged 8-10 years). The programme consisted of ten 40-minute sessions which were incorporated into the children’s weekly schedule and delivered as part of their normal school day. A combination of standardised and bespoke outcome measures were used to evaluate the effectiveness of the group pre- and post-intervention. Following the intervention, well-being was shown to improve across several areas of functioning; children’s knowledge regarding emotional literacy and thought/feeling/behaviour discrimination improved; there was improvement in the children’s tendency to agree with the characters who were jumping to conclusions; and both children and teachers reported learning key CBT strategies and spoke positively about their experience of CUES-Ed. The limitations and reflections from this case are also discussed.

## CONTENTS

---

<b>1. LITERATURE REVIEW</b>	<b>87</b>
1.1. Mental Health Needs in Children	87
1.2. Universal Classroom-based Interventions	87
<b>2. CASE DESCRIPTION</b>	<b>90</b>
2.1. Setting	90
2.2. The Children	90
<b>3. INTERVENTION</b>	<b>92</b>
3.1. CUES-Ed	92
3.2. Session Protocol and Adaptions	93
3.3. Facilitators	95
3.4. Outcome Measures	95
<b>4. RESULTS</b>	<b>98</b>
4.1. Aim 1 (to teach basic CBT strategies to manage difficult feelings)	98
4.2. Aim 2 (to improve general well-being)	99
4.3. Aims 3 & 4 (to improve emotional well-being and behaviour)	100
4.4. Aim 5 (to improve ability to look after own basic well-being)	100
4.5. Aims 6 & 7 (to improve emotional literacy and repertoire of coping strategies)	101
4.6. Aim 8 (to increase mental health awareness)	102
4.7. Aim 9 (to normalise responses to confusing or difficult situations)	102
<b>5. DISCUSSION</b>	<b>103</b>
5.1. Summary and Limitations	103
5.2. Reflections	105
<b>6. REFERENCES</b>	<b>107</b>
<b>7. APPENDICES</b>	<b>111</b>
7.1. Table of CBT adaptions extracted from the literature	111

## 1. LITERATURE REVIEW

---

### 1.1. Mental Health Needs in Children

In 2015, the Department of Health (DoH) published the “Future in Mind” report which highlighted the need for an appropriate system to adequately support the emotional wellbeing and mental health of children, young people, and their families. One in ten children aged 5-16 years were found to have clinically significant mental health problems, where 5.8% of children were diagnosed with conduct disorders, 3.7% emotional disorders, and 1.5% hyperkinetic disorders (Meltzer, Gatward, Corbin, Goodman & Ford, 2003). Unresolved emotional difficulties have been found to result in lower educational attainment and behaviours which can be further detrimental to mental and physical health (including, smoking, alcohol and drug use, and risky sexual behaviour) (DoH, 2015).

Several government publications have arisen which aim to review and improve child and adolescent mental health, including: a green paper (“Transforming Children and Young People’s Mental Health Provision”; DoH, 2017); a mental health strategy (“No Health Without Mental Health”; DoH, 2011); and a policy paper (“Prevention concordat for Better Mental Health”; Public Health England, 2018). A common trend throughout all these documents is the recommendation for the prevention and early intervention of mental health. Such programmes (for both children who have and have not experienced mental health difficulties) could be effective in reducing the incidence and prevalence of mental health disorders over time and the subsequent costs, as well as the likelihood of young people entering a state of crisis or requiring expensive interventions in adulthood (DoH, 2015; McCrone, Dhanasiri, Patel, Knapp & Lawton-Smith, 2008). However, the National Child and Adolescent Mental Health (CAMHS) Support Service (2011) recognised that commissioners can find such programmes challenging, due to the limits on resources and the cost-benefit being difficult to demonstrate in the short-term.

### 1.2. Universal Classroom-based Interventions

Schools play a key part in a child’s development, particularly the areas that also appear to be affected by mental health difficulties. For instance, peer relationships and social interactions, academic attainment and cognitive progress, emotional control and behavioural expectations, and physical and moral development (Fazel, Hoagwood, Stephan & Ford, 2014). Anticich and colleagues (Anticich, Barrett, Gillies & Silverman, 2012) outlined the benefits of delivering preventative interventions within the school environment. They highlight how school is the

second most frequented setting by children, thus providing a natural resource for children/families to develop coping strategies in a real-world setting. They additionally argue that schools are easily and freely accessible, which would overcome the difficulties in accessing community based mental health services and is economically viable, as well as how a school setting would further normalise mental health (reducing stigma) and allow a context for children to adequately practice their skills.

A three-tiered model exists for carrying out preventative interventions in schools, with the three components being: universal, selective/targeted, and indicated interventions (Fazel et al., 2014). Universal programmes are aimed at a whole population of children (e.g. whole classrooms or schools); targeted programmes are delivered only to children who are at an increased risk of a disorder compared to the average child (e.g. children with parents who have mental illness); and indicated programmes are indicated for those already exhibiting signs of mental health difficulties or have established diagnoses (Wells, Barlow, & Stewart-Brown, 2003).

It is universal programmes that would arguably allow greater proportions of children to access and benefit from early intervention. Benefits of universal interventions include the minimisation of stigma (as particular individuals at risk are not being identified), the removal of screening resources, and being inclusive of children who are not currently at risk but would go on to develop symptoms in future (Werner-Seidler, Perry, Caelear, Newby & Christensen, 2017). Universal approaches can also appeal to schools due to not being intrusive, incurring relatively low costs, and being easily incorporated into the curriculum (Fazel et al., 2014). They have further been found to be more effective compared to selective and indicated programmes for anxiety (Neil & Christensen, 2009), however for depression, the most efficacious programme design was indicated programmes (Caelear & Christensen, 2010).

A recent review was carried out by Werner-Seidler and colleagues (2017), which aimed to comprehensively evaluate all the published Randomised-Control Trials (RCTs) of school-based prevention programmes for depression and/or anxiety in children. This review included 81 studies yielding over 30,000 young people. The findings showed that these preventative interventions had a small beneficial effect on mental health symptoms compared to a control group, which remained modest over long-term follow up (i.e. one year). The authors note that despite the small effect size found, in the context of preventative interventions (as opposed to treatment), such findings would still have great implications in preventing the onset of disorders in children and is therefore still ecologically valuable. Similarly, the DoH (2017) undertook a comprehensive systematic review of general mental health in young children, however found that universal prevention programmes had limited evidence for the prevention of depression

and anxiety. Even though, this review was not exclusive to RCT studies (i.e. the “gold-standard” research design). Positively, the authors suggested that universal interventions create “herd effects” which lead to students who are not specifically targeted having better outcomes. Such school-based programmes have also been demonstrated to be more effective when delivered by health professionals compared to trained teachers (Stallard et al., 2014).

“CUES-Ed” is an example of a universal, mental health prevention package that is delivered to mainstream primary schools within the boroughs of London. It is yet to be empirically evaluated, although anecdotally it has obtained positive outcomes across a variety of schools, where children identified as having special educational needs have shown to benefit the most. This case study outlines the delivery of an adapted version of the CUES-Ed programme for children with Social, Emotional, and Mental Health (SEMH) needs.



## 2. CASE DESCRIPTION

---

### 2.1. Setting

The CUES-Ed team were initially approached by a specialist school for children with Social, Emotional, and Mental Health (SEMH) needs in Central London to provide the CUES-Ed programme to its pupils, who would otherwise not have access to such an intervention because they were not in a mainstream school. The CUES-Ed team accepted the request on the basis that the programme aims to provide *all* children with the knowledge and skills to manage difficult emotions. The Headteacher acknowledged the differences in learning needs between its SEMH children and those from mainstream schools, but understandably was keen for as much parity in resources as possible.

The school is a specialist establishment that meets the educational needs of children with SEMH difficulties. Such SEMH difficulties might present as problems with mood (anxiety or depression) or conduct (oppositional problems and more severe conduct problems including aggression), as well as disorders including: Attention Deficit Disorder (ADD), Attention Deficit Hyperactive Disorder (ADHD), attachment disorder, autism or pervasive developmental disorder, an anxiety disorder, a disruptive disorder or, rarely, schizophrenia or bipolar disorder (Department of Education, 2016). Children with SEMH needs may present as being withdrawn and isolated or with behaviour that challenges others (e.g. physical/verbal aggression, impulsive, confrontational and sexualised behaviour), and are typically those in or on the edge of care, endured significant adversities and trauma, and/or suffer from attachment disorders (Lenehan & Geraghty, 2017). It is not surprising then that two thirds of these children become permanently excluded from mainstream schools and subsequently require placements in special schools like this one (Cole, 2015).

Children with SEMH needs experience many individual, familial, and environmental risk factors for mental health disorders and consequently lack the protective resilience factors known to be essential for good mental health, including problem-solving, communication skills, and positive attitude (Cole, 2015). Therefore, it makes logical sense that CUES-Ed is made available and accessible to these children, as they are arguably the ones who need it most.

### 2.2. The Children

Demographically, the SEMH school is designed as a mixed school for all genders and ethnicities. The adapted CUES-Ed programme was delivered to two separate classes within the school. Each

class was of mixed ability, where children were either in school years 4 or 5 (aged 8-10 years). Across the two classes there were 14 children, seven from each class, only one of whom was female. There were no inclusion/exclusion criteria, and the group was placed onto the class' weekly schedule and delivered as part of their usual school day.

Due to the school's confidentiality procedures, specific details regarding the children's mental health and social circumstances could not be disclosed. Broadly speaking, the group represented a diverse range of ethnic heritages, and due to being considered children with SEMH difficulties, all had a statement of special educational needs. Typical presentations in the classes were: ASD, ADHD, speech and language difficulties, conduct problems, and learning difficulties. There was at least one child subject to a care order, one child open to and being assessed by social services, and others who may have previously been known to Local Authorities.

### 3. INTERVENTION

---

Universal approaches are recommended within the National Institute for Health and Clinical Excellence (NICE) guidelines (2008) as an appropriate intervention for primary education. Subsequently, this model was used to form the basis of the group intervention. NICE also suggest that universal programmes should additionally target parents (e.g. by providing information or offering small, group-based programmes), and delivered by appropriately trained teachers and practitioners.

#### 3.1. CUES-Ed

“CUES-Ed” was a whole-class psychoeducation programme designed to teach primary school children (aged 7-10 years) how to notice the “cues” for when things aren’t right, as well as practical coping strategies to manage difficult thoughts and feelings, both at home and at school (South London and Maudsley NHS Trust, 2018). It was developed in response to feedback from children in CAMHS settings and a Cognitive Behavioural Therapy (CBT) research trial, who reported that they would have found the therapy more useful should they have been taught the concepts earlier (Maddox et al., 2013). They also reported appreciating the normalising approach of therapy as they had previously witnessed high levels of stigma. The main aims were:

- Aim 1: To teach all children basic CBT strategies to manage difficult feelings
- Aim 2: To improve general well-being
- Aim 3: To improve emotional well-being
- Aim 4: To improve behaviour
- Aim 5: To improve children’s ability to look after their basic well-being
- Aim 6: To improve emotional literacy
- Aim 7: To increase repertoire of coping strategies
- Aim 8: To increase mental health awareness
- Aim 9: To normalise responses to confusing or difficult situations

This traditional CUES-Ed programme consisted of six one-hour interactive sessions, as well as two additional sessions for the administration of pre- and post-group outcome measures. Teachers and parents were further encouraged to be involved in sessions (as per NICE guidelines), where parent handouts were sent out following each session and mid-week emails

were sent to teachers to maintain/encourage engagement. The class were also provided with a poster after each session to summarise the key learning points.

### 3.2. Session Protocol and Adaptions

Adaptions for the SEMH school were informed by the literature, which is summarised in Appendix 7.1. Of note, the traditional CUES-Ed programme already included a number of these adaptions, as it was intentionally designed to be inclusive, where even in typical classrooms SEMH needs and learning difficulties are prevalent. These practices include: presenting information both visually and verbally, using simple language, involving parent and teachers in the programme, repeating key concepts, incorporating activities and setting homework tasks, and modelling behaviour. Additionally, SEMH schools adopt several of the recommendations in Appendix 7.1. as part of their standard practice (e.g. token reinforcement, incorporating extra staff support, including CUES-Ed on the weekly scheduling, etc.).

Subsequently, the adapted CUES-Ed programme followed the same practices and aims that were in the traditional programme, although the main differences were that it included an additional two sessions (10 sessions in total), and sessions lasted approximately 40 minutes in length (compared to an hour). The content of the original sessions was significantly reduced to fit with this shorter session time, and the language was simplified further. New activities were also incorporated to increase practical skills and minimise extended periods of didactic teaching. All sessions (apart from the introductory session) began with a review on the previous session's learning and tasks, as well as including an agenda for the current session. Ground rules were set at the beginning and visual aids were used on each slide to communicate whether the children's task was to listen, put their hand up and engage in discussion, or take part in an activity.

As with the original CUES-Ed programme, all children were provided with a CUES-Ed work book which was designed to be used both within and outside of sessions, received session posters and parent handouts, and teacher emails were sent between sessions. A summary of the adapted CUES-Ed sessions can be found in Table 7.

**Table 7: To show the adapted CUES-Ed session structure along with each session's key learning themes/aims.**

Session	Name	Key Themes and Learning
<b>PRE</b>	Introduction and Pre-Group Outcomes	<ul style="list-style-type: none"> <li>- What CUES-Ed is.</li> <li>- Introduction to the brain.</li> </ul>
<b>1</b>	Looking After Ourselves, Part I	<ul style="list-style-type: none"> <li>- Introduction to the programme characters (Ed and Chloe).</li> <li>- Individual difference.</li> <li>- Brain controls our emotions.</li> <li>- Importance of eating well and being active.</li> <li>- Linking physical and emotional well-being.</li> </ul>
<b>2</b>	Looking After Ourselves, Part II	<ul style="list-style-type: none"> <li>- Importance of sleeping well and relaxation.</li> <li>- Linking physical and emotional well-being.</li> </ul>
<b>3</b>	Noticing How We Feel, Part I	<ul style="list-style-type: none"> <li>- Emotional literacy (i.e. the clues to how we or others are feeling).</li> <li>- Understanding that feelings can change quickly.</li> <li>- Specific focus on behaviours and body sensations relating to anger.</li> <li>- Promote coping strategy of noticing own feelings</li> </ul>
<b>4</b>	Noticing How We Feel, Part II	<ul style="list-style-type: none"> <li>- Specific focus on physiological response to anxiety and the differences/similarities with anger.</li> <li>- Fight or flight response as reason for these symptoms, as well as real vs. false alarms.</li> <li>- Breathing Strategy to manage physiological symptoms</li> </ul>
<b>5</b>	Thoughts, Feelings, and Behaviours	<ul style="list-style-type: none"> <li>- Differentiating thoughts, feelings, and behaviour.</li> <li>- Introduction to the concept that thoughts, feelings and behaviour are all linked.</li> <li>- Helping children to identify thoughts</li> <li>- Teach concept that thoughts are just thoughts.</li> <li>- Promote coping strategy to talk to trusted others about feelings</li> </ul>
<b>6</b>	Finding Helpful Thoughts	<ul style="list-style-type: none"> <li>- Behavioural strategy to inhibit a response that might result in a problem</li> <li>- Exploring the difference between helpful and unhelpful thoughts, and how they matter for what we do.</li> <li>- Introducing cognitive strategies such as noticing cognitive biases.</li> <li>- Introduce positive self-talk</li> </ul>

Session	Name	Key Themes and Learning
7	Our Amazing Brains	<ul style="list-style-type: none"> <li>- Introducing normalising explanations to confusing or unusual situations.</li> <li>- Exploring the implications of jumping to conclusions.</li> <li>- Cognitive strategies in response to jumping to conclusions.</li> </ul>
8	Finding Out What I Can	<ul style="list-style-type: none"> <li>- Building on the existing strategies.</li> <li>- Learning about the processes of worry and anger rumination.</li> <li>- Introducing mindfulness and noticing the environment.</li> </ul>
POST	Review, Certificates, and Post-Group Outcomes	<ul style="list-style-type: none"> <li>- Celebration of what children have learnt.</li> </ul>

### 3.3. Facilitators

The group was facilitated by one qualified and one trainee clinical psychologist. As part of the programme, at least one teacher and/or teaching assistant was required to remain in the classroom throughout the delivery of the group. This was to support the facilitators in the behavioural management of the classroom and to comply with regulations specific to the school. Additionally, it was integral for teachers to remain present throughout to learn the CUES-Ed techniques and philosophy themselves, so they can be applied outside of sessions. This allowed for the children to have greater opportunities to consolidate the skills.

### 3.4. Outcome Measures

To effectively evaluate the effectiveness of the adapted CUES-Ed programme, various outcome measures were used both pre- and post-intervention. These measures were collated into a “CUES-Ed Quiz” booklet and presented in a child-friendly manner with various visual cues. Both standardised and unstandardized, bespoke measures were used. Those which were unstandardized were developed idiosyncratically by the CUES-Ed authors to measure the specific aims and learning outcomes of the CUES-Ed programme. All measures are outlined in Table 8 below, along with the corresponding CUES-Ed aim.

Additionally, child, teacher, and parent feedback was used to further evaluate the effectiveness of the programme, and specifically used to assess whether Aim 1 (to teach basic CBT strategies to manage difficult feelings) was met.

**Table 8: To show the outcome measures used to evaluate the adapted CUES-Ed programme, and which CUES-Ed aim it maps on to.**

Measure and Reference	Corresponding CUES-Ed Aim	Description
Standardised Outcome Measures		
<b>Child Outcome Rating Scale (CORS)</b> (Duncan, Sparks, Miller, Bohanske & Claud, 2006)	Aim 2: Improve general well-being	The CORS is a 4-item self-report measure for children (6-12 years), assessing four areas of functioning/wellbeing (individually, interpersonally, socially, and overall). Children were asked to indicate on a 0-10 likert scale how well they are doing in each of the four areas, with the zero-end of the scale being marked by a sad face and representing extremely negative wellbeing, and the opposite end (10) by a happy face and representing excellent wellbeing. The CORS has demonstrated good validity, reliability, and feasibility.
<b>Me and My Feelings (M&amp;MF)</b> (Deighton et al., 2013)	Aim 3: Improve emotional well-being	The M&MF is a 16-item self-report questionnaire for children as young as 8 years old. It is comprised of two subscales, each measuring either emotional (10-items) and behavioural difficulties (6-items). Children are required to read a statement and respond on a 3-point likert scale for how frequently this is true for them (“never”, “sometimes” or “always”). One item is reverse-scored, and total scores can range from 0-32, with higher scores representing greater difficulties. Scores of 12 and above on the emotional subscale, and seven or above on the behavioural subscale indicate clinically significant difficulties. This scale has demonstrated good levels of internal consistency and construct validity.
	Aim 4: Improve behaviour	
Bespoke Outcome Measures		
<b>Adapted CORS (CUES-Ed)</b>	Aim 5: Improve children’s ability to look after their basic well-being	In addition to the original four CORS items, the CUES-Ed team added four further items using the same original format. Specifically, these additional items asked children “how are things going with...” eating healthy, being active, sleeping, and relaxing. These were included as they are areas specifically taught in the CUES-Ed package as they are known to positively impact on emotional well-being.
<b>Emotional Literacy (CUES-Ed)</b>	Aim 6: Improve emotional literacy	Children are provided with a list of possible clues to how they or someone else might be feeling amongst distractor items. Children tick items they believe are clues. A total score is calculated (correct items – incorrect).
<b>Thought, Feeling, and Behaviour Discrimination (CUES-Ed)</b>	Aim 6: Improve emotional literacy	To assess the child’s ability to discriminate between thoughts, feelings, and behaviours, two scenarios were provided involving the two CUES-Ed characters. Children are required to differentiate appropriate thoughts, feelings and behaviours from distractor items. A total score is calculated allowing 1 mark for every correct answer giving a total out of 6.

Measure and Reference	Corresponding CUES-Ed Aim	Description
<b>Coping and Resilience</b> (CUES-Ed)	Aim 7: Increase repertoire of coping strategies	Children were provided with two scenarios and asked to identify which strategies from a list would be the most helpful in the given situation. Children must differentiate helpful from unhelpful strategies and are given a total score (number of helpful – unhelpful strategies).
<b>Stigma</b> (CUES-Ed)	Aim 8: Increase mental health awareness	Children's beliefs about showing and talking about emotions was assessed using three items. The first two ask the child to rate on a 4-point likert attitudes towards expression of emotion, firstly for themselves, and then for others. The scale ranged from "it is never okay" to "it is mostly okay". The final item asked the child to select words from a list to finish of the sentence "When someone shows to talks about their feelings when they are sad, worried, or angry, I think that it is...". Responses included a combination of neutral/non-stigmatising appraisals and stigmatising appraisals. A total score was calculated (number of non-stigmatising items – stigmatising items).
<b>Normalising Unusual Experiences</b> (CUES-Ed)	Aim 9: Normalise responses to confusing or difficult situations	Children were asked to tick one of six possible explanations to a CUES-Ed character experiencing a potential confusing or unusual experience. The answers included normalising and negative appraisals. Children's scores were then categorised as either normalising appraisal or negative appraisal.
<b>Jumping to Conclusions</b> (CUES-Ed)	Aim 9: Normalise responses to confusing or difficult situations	Children were provided with two scenarios in which the characters jumped to a negative conclusion. Children rated on a likert scale the characters' appraisals ranging from 0-5, where 0 represented that the character could not know for sure, and 5 represented they were definitely right. The children's level of certainty was evaluated with children who scored 5 on either or both measures being more certain in their responses.



## 4. RESULTS

---

The results of the adapted CUES-Ed intervention are presented below and organised according to the CUES-Ed aims.

Unfortunately, due to the nature of the school and client group, outcome measures were unable to be obtained by all children at both time-points. This was due to either refusal to participate, leaving the classroom during this time, or being absent from school. Therefore, any references to outcomes refer to only those children who provided both pre- and post-intervention data ( $N = 6$ ). Qualitative feedback is only representative of nine of the children and one of the teachers. No parents completed feedback.

### 4.1. Aim 1 (to teach basic CBT strategies to manage difficult feelings)

Feedback from teachers ( $N = 1$ ) and children ( $N = 9$ ) evidenced that the children had learnt several CBT strategies to help them manage difficult feelings.

In their comments, two children specifically referenced breathing techniques as a behavioural strategy used to manage their emotions. Two children also referred to cognitive strategies, such as identifying thoughts and noticing cognitive biases. Other aspects of the CUES-Ed programme that children reported learning about included thinking positively ( $N = 1$ ), about other people's feelings ( $N = 1$ ), and facts about the brain ( $N = 2$ ). All children reported finding CUES-Ed helpful and fun, and that they understood the content at least some of the time. Children were asked to record how CUES-Ed had helped them, some responses can be found below:

*"It's made me be a much more healthy person."*

*"Helped me about stress."*

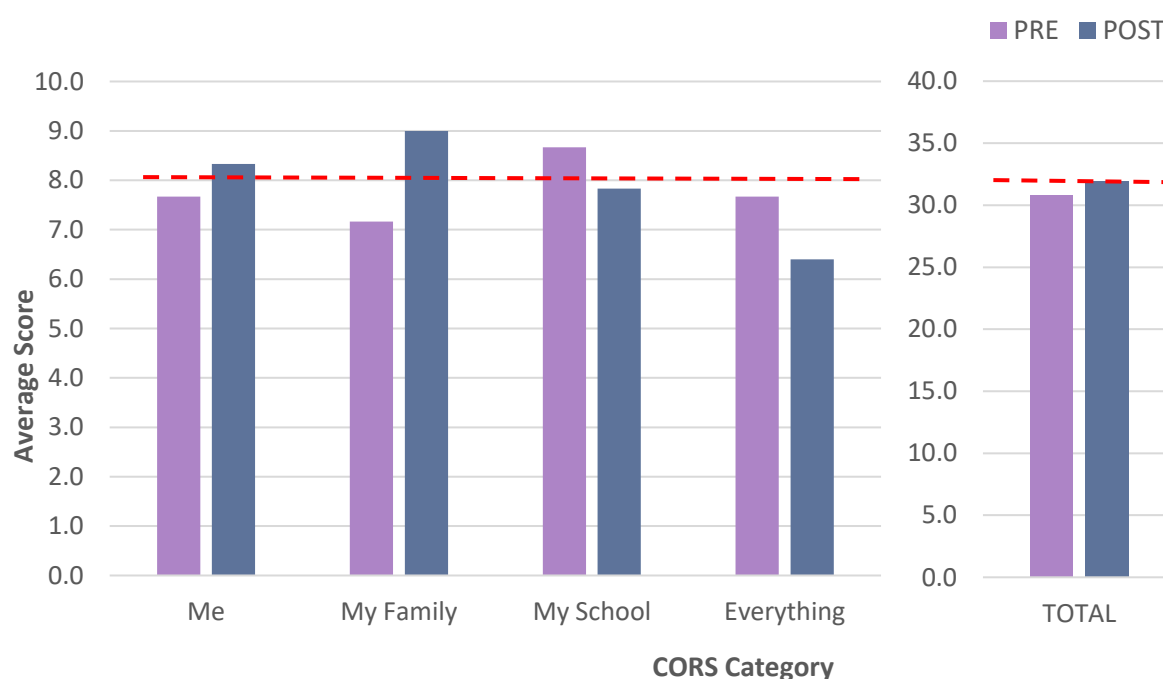
*"With my anger."*

The teacher agreed with the children in finding the programme helpful for the class and described themselves learning a variety of cognitive and behavioural techniques which they planned to continue to use, such as positive self-talk and anger management techniques. The teacher noted that the CUES-Ed programme had helped them to consider the way in which they communicate with the children about difficult feelings, as well as how to subsequently support them with these. The teacher was specifically able to provide examples of the children utilising

the CUES-Ed language outside of sessions. For example, one child disclosed a normalising explanation for an unusual experience using language taught on the programme. Another recalled all the facts they had learnt about the brain. This teacher also reported improvements in one specific child's behaviour following CUES-Ed, particularly in helping him to notice when he was getting angry and so was more likely to take appropriate action. Suggested improvements included incorporating even more practical activities into this adapted CUES-Ed, using fewer handouts/written tasks, and increasing opportunities for repetition and practice.

## 4.2. Aim 2 (to improve general well-being)

Children's wellbeing was measured using the original version of the CORS. There is a clinical cut-off set at 8 per item (32 for the total score), where higher scores represent greater wellbeing. Average total scores on the CORS improved following CUES-Ed, from being clinically significant to reaching the expected range (Figure 7). Children's perspectives on their own and family well-being improved to the expected level following CUES-Ed, although "school" and "everything" ratings decreased.



**Figure 7: Bar graph to show mean average pre- and post-intervention scores on the CORS categories and total score, as well as the clinical cut-off boundaries (red dashed line).**

### 4.3. Aims 3 & 4 (to improve emotional well-being and behaviour)

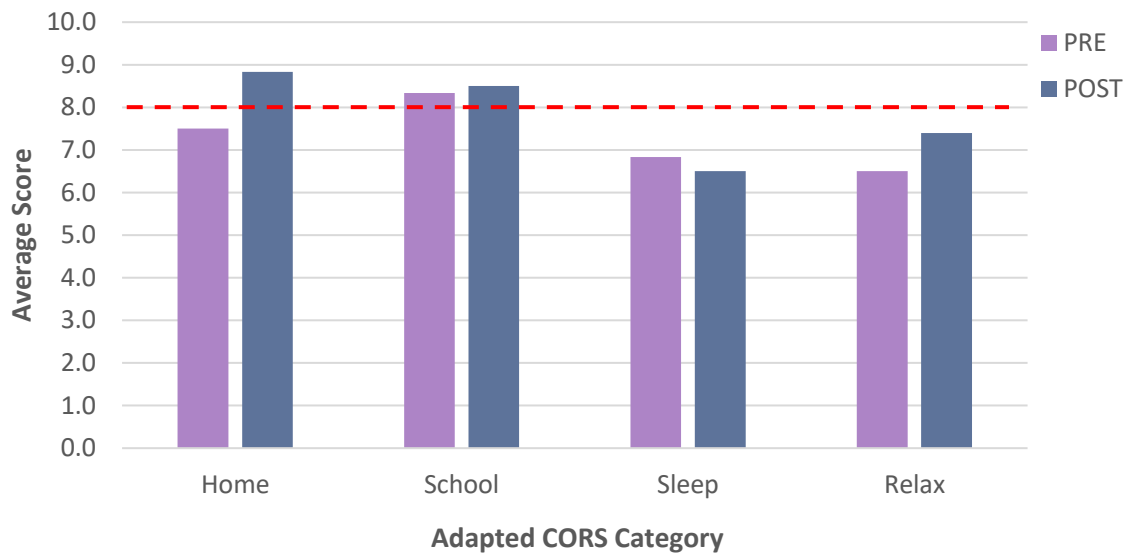
Children were asked to indicate emotional and behavioural difficulties, where lower scores are desirable as they represent fewer difficulties. Positively, children's self-report both pre- and post-intervention were below the clinical cut-off for both the emotional and behaviour scales (Figure 8). Participating in CUES-Ed resulted in no change on the emotional subscale, and a decrease of 0.2 on the behavioural subscale (and subsequently total score).



**Figure 8: Bar graph to show mean average pre- and post-intervention scores on the M&MF. The red dashed line represents the clinical cut-offs on each subscale.**

### 4.4. Aim 5 (to improve ability to look after own basic well-being)

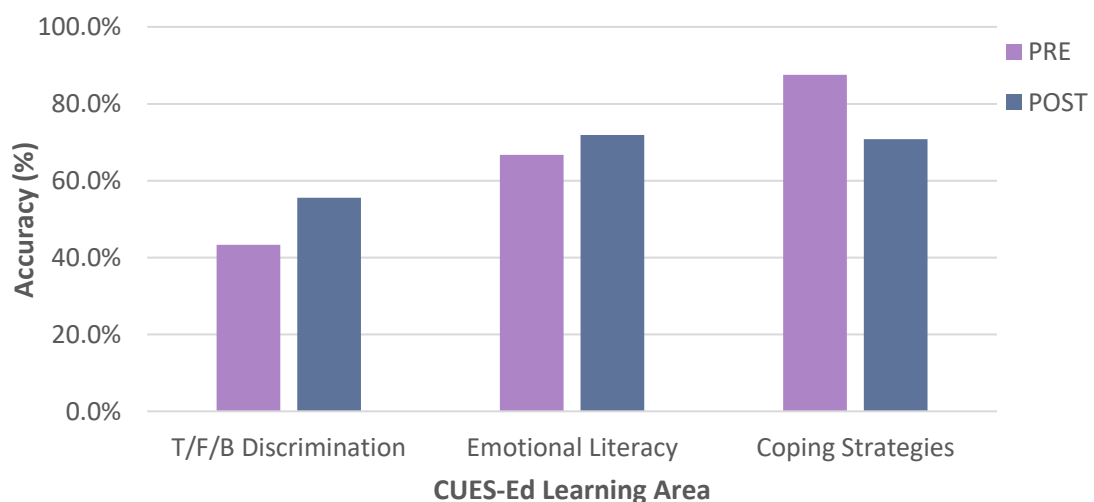
The CUES-Ed adapted CORS was used to measure this aim, results can be found in Figure 9. Since completing the CUES-Ed programme, children showed improvements on all categories, except for sleep. Only the active category reached the expected level post-intervention. The 'cut off' for this measure was 8 for consistency with the CORS.



**Figure 9: Bar graph to show mean average pre- and post-intervention scores on the CUES-Ed adapted CORS, as well as the clinical cut-off boundary (red dashed line).**

#### 4.5. Aims 6 & 7 (to improve emotional literacy and repertoire of coping strategies)

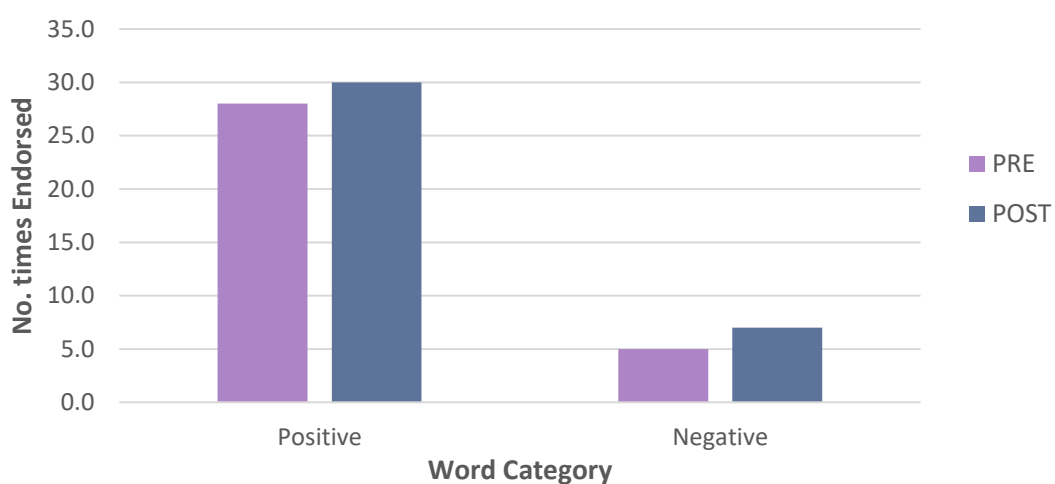
Outcomes measuring accuracy on tasks of emotional literacy, thought/feeling/behaviour discrimination, and identification of adaptive coping strategies pre- and post-CUES-Ed are shown in Figure 10. Children's thought/feeling/behaviour discrimination and emotional literacy improved as a result of the programme. However, on average, their ability to identifying appropriate coping strategies decreased.



**Figure 10: Bar graph to show mean accuracy (%) on the CUES-Ed key learning areas pre- and post-intervention.**

#### 4.6. Aim 8 (to increase mental health awareness)

CUES-Ed additionally aimed to reduce the stigma associated with difficult emotions. Before CUES-Ed, 83% of children stated that it was either “sometimes” or “mostly okay” to show and talk about *their own* feelings, and 100% believed this *in relation to others*. Against expectation, this decreased to 67% and 83% respectively following CUES-Ed. Children were further asked to select multiple words from a list to describe someone who talks about/shows their feelings. The results are displayed in Figure 11. Overall, negative words were endorsed considerably less, both pre- and post-CUES-Ed compared to positive words. Post-intervention, the number of times both positive and negative words were endorsed increase by two.



**Figure 11: Bar graph to show the number of times positive/negative words were endorsed, both pre- and post-intervention.**

#### 4.7. Aim 9 (to normalise responses to confusing or difficult situations)

Regarding the normalisation of unusual experiences, children were asked before and after CUES-Ed to select one explanation for why a character could be having an unusual experience. Prior to the CUES-Ed intervention, 100% of children chose a normalising rationale, which was maintained after the intervention. Finally, in two separate scenarios, children were asked to decide whether the conclusions drawn by the characters (who were jumping to conclusions) were reasonable or not. On average, children were less likely to agree with the characters jumping to conclusions post-intervention compared to pre-intervention (reduction from 1.9 to 1.5 – higher scores (range = 0-5) represent greater character agreement). This suggested the children showed less certainty in their appraisals, an expected outcome of the programme.

## 5. DISCUSSION

---

### 5.1. Summary and Limitations

Two classes ( $N = 14$ ) in a school for children with SEMH needs participated in an adapted version of a universal psycho-education package called “CUES-Ed”. CUES-Ed was designed as an early-intervention programme that aimed to improve emotional well-being and resilience, in addition to reducing mental health-related stigma and normalising unusual experiences. The primary adaptations included providing an increased number of shorter duration sessions, incorporating more practical tasks to demonstrate concepts, and simplifying the delivery and language.

Following engagement in the programme, wellbeing improved in half of the areas reported on the original CORS and in three of the four categories on the adapted CORS. It was the sleep domain where a reduction in scores were observed, but it could be argued that, as a child, this is more difficult to change due to systemic issues. For example, at home the child may have less control over sleeping conditions compared to those associated with relaxing and being active, especially in the context of their social difficulties (e.g. may share a room with older siblings, have noisy households, etc.). Moreover, no change was observed in self-reported emotional difficulties but there was a slight decrease in behavioural difficulties post-CUES-Ed. CUES-Ed appeared to increase the children’s emotional literacy and understanding of thoughts, feelings, and behaviours, as evidenced by increased scores post-intervention. Conversely, the post-group outcomes demonstrated a reduction in accuracy for identifying adaptive coping strategies, despite the use of strategies being indicated in child and teacher qualitative reports. When prompted (e.g. by teachers, CUES-Ed staff, or questionnaires), the children were able to access this information. This could suggest that more repetition and generalisation was required to embed such skills and generalise them out of the classroom context, so that they are more likely to recall and utilise such strategies independently. This could have been achieved through increasing the contact and involvement from teachers and parents/carers, which is known to be an effective but challenging component of child programmes (Hirshfeld-Becker & Biederman, 2002).

Positively, both pre- and post-intervention, the children endorsed more positive words with showing/talking about emotions compared to negative ones. Nevertheless, there was also an increase in the selection of negative descriptors. Finally, there were improvements in scores post-intervention which measured certainty in cognitive appraisal (jumping to conclusions) and all children provided a normalising rationale to two confusing/difficult scenarios. Taken together, the adapted CUES-Ed programme was able to meet a number of the aims that it set

out to achieve, namely that children had: learnt basic CBT strategies, showed improvements in most areas of general well-being, showed improvements in behaviour, learnt emotional literacy skills, increased their mental health awareness, and were more likely to find confusing or difficult situations normal. Improvements in emotional well-being and repertoire of coping strategies was not observed as hoped for this particular group of children. This demonstrates that, on the whole, adapting early intervention packages for children with SEMH needs (where the children arguably have greater mental health difficulties) was effective in improving the children's resilience. Therefore, CUES-Ed and other similar psycho-education programmes should continue to make the content accessible for children with SEMH needs, and actively encourage specialist schools to incorporate them into their learning programme.

Subsequently, these results support NICE (2008) and government recommendations (DoH, 2011, 2015, 2017; Public Health England, 2018) to incorporate more universal and early-intervention programmes to target mental health in schools. To increase the efficacy of the findings, should such an evaluation be conducted again, it would be worthwhile to also collect longitudinal data to identify whether any observed changes remain over time. It would also be beneficial to collect more informant data (e.g. teacher/carer report of child) to increase the reliability of the outcomes.

Unfortunately, attendance was sporadic and inconsistent across the classes, which may account for why not all CUES-Ed aims were met. Children may have been absent for the whole day or removed before/part-way through the class due to various reasons (e.g. conflicts with others, alternative educational activities). Additionally, engagement with the session and associated tasks varied, where occasionally children would decline to participate, which was the case for the pre-intervention outcome measures. This meant that pre- and post-outcome data was not reflective of all the children who participated in CUES-Ed and the sample size was small. Subsequently, when negative changes in scores were observed against the expected direction, these were often accounted for by only one individual.

Moreover, it is important to consider the presence of conduct and behavioural difficulties in the client group. Therefore, it was not unusual for particular children to test boundaries and be intentionally oppositional (especially when the sessions were ending), and it is likely that the observations in post-outcome scores against expectations correspond with these individuals. Consequently, the findings of this adapted CUES-Ed programme should be interpreted with caution and may be an underestimation of the group's true effects, should children have been able to fully attend and engage in sessions. However, this could also reflect that the CUES-Ed programme was not the best suited for children with conduct and oppositional difficulties,

regardless of the adaptations made. The NICE guidelines (2013) suggest that for children aged 9-14 years with conduct disorder, cognitive-behavioural interventions should primarily focus on problem-solving and consist of 10-18 2-hour weekly sessions. Therefore, the current CUES-Ed programme may have not fully met the needs of these children by providing an intervention which was not long or intensive enough, or that focused on the wrong topics. Nonetheless, the teacher reported an increased awareness of the children's needs and learned new skills to manage them, which would have indirectly benefitted the children.

An additional adaptation to consider in the CUES-Ed programme for SEMH schools, is whether the philosophy should be embedded in the whole school, rather than individual classes. In their green paper, the government suggest that "a whole school approach, with commitment from senior leadership and supported by external expertise, is essential to the success of schools in tackling mental health" (p. 5; DoH, 2017). Such SEMH schools typically have minimal students and classes, and therefore would not entail the same burden to facilitate compared to mainstream primary schools. This would also ensure that the children were consistently exposed to the key CUES-Ed principles/language and encouraged to draw on the coping strategies during conflicts or other difficult emotional and behavioural experiences. Parents would then also be more exposed to this way of working, which would all contribute to the generalisation of skills and knowledge in different contexts, and therefore children would have even more opportunities to benefit from the CUES-Ed programme.

In the current school, the Headteacher noted the differences in engagement of the two classes and attributed this to variations in teacher engagement (as one teacher had expressed uncertainty in the programme, believing that it would work better as an individual intervention). Teacher investment/engagement was not measured in this case study but may be an important factor to consider in future. A greater level of induction before the programme may have been a way to overcome this.

## 5.2. Reflections

The delivery of the adapted CUES-Ed programme came with several benefits and challenges. I was not used to working with a group of vulnerable and behaviourally challenging children, my experience was only with individual clients. I found it very rewarding to observe how the children slowly warmed up and welcomed us over time. It demonstrated to me the power of adopting a clear boundaried, consistent, and patient approach, as you are essentially a stranger invading their safe space. I also learnt the benefits of slowing down my pace and being more considerate in my language, as well as how difficult I found this to do in practice. By comparing my experience



of CUES-Ed in mainstream and SEMH schools, it was evident why such a programme required adapting, especially as these children are the ones who could potentially benefit the most. I believe one of the most influential components of the group was the inclusion of more practical tasks to illustrate key concepts, as it overcame several cognitive difficulties that are required when listening/observing. I was also surprised to notice how well these children were able to access the cognitive components of the group and were less interested in the behavioural strategies – where I would have predicted this to be the other way around.

I found myself feeling frustrated at several points across the group, mainly directed at particular children and staff. I was aware at how much more the other children got out of the sessions when a certain child was not present. This individual was often oppositional to the session and would purposely attempt to be disruptive. It made me question whether he should be removed from the class, but I was aware that this was associated with my desire to provide a good and meaningful intervention. Despite this child's behaviour, they were still likely to have processed some key messages and benefitted in some way. Regarding the staff, understandably they were more experienced in working with this client group and were good at managing class dynamics. However, I found one of the teachers to not always be open to our way of working due to beliefs that it would not be effective, and therefore did not engage in or promote CUES-Ed outside of the sessions. I also found the school in general to reflect the client group in some ways – somewhat chaotic – as both professionals and children moved in and out of classrooms without explanation, where there was even a time when we were left unsupervised with the children. As a competent adult, this made me feel uncontained, confused, and even neglected, where it left me wondering what impact this disruption had on the children, especially as this may be mirrored in their home lives. I genuinely believe that a whole-school approach could have been more influential for this client group and would recommend this in the future.

## 6. REFERENCES

---

- Antich, S. A. J., Barrett, P. M., Gillies, R. & Silverman, W. (2012). Recent Advances in Intervention for Early Childhood Anxiety. *Australian Journal of Guidance and Counselling*, 22, 157-172.
- Calear, A. L., & Christensen, H. (2010). Systematic review of school-based prevention and early intervention programs for depression. *Journal of Adolescence*, 33, 429–438.
- Cole, T. (2015). *Mental Health Difficulties and Children at Risk of Exclusion from Schools in England: A review from an educational perspective of policy, practice and research, 1997 to 2015*. Retrieved from <http://www.education.ox.ac.uk/wordpress/wp-content/uploads/2015/02/MENTAL-HEALTH-AND-EXCLUSION-FINAL-DIGITAL-13-06-15.pdf>.
- Council for the Curriculum, Examinations and Assessment (CCEA). (2016). *Young Learners and Social, Emotional and Behavioural Difficulties (SEBD)*. Retrieved from [http://ccea.org.uk/sites/default/files/docs/curriculum/guidelines\\_general\\_strategies/Young\\_Learners\\_and\\_Social\\_Emotional\\_Behavioural\\_Difficulties.pdf](http://ccea.org.uk/sites/default/files/docs/curriculum/guidelines_general_strategies/Young_Learners_and_Social_Emotional_Behavioural_Difficulties.pdf).
- Deighton, J., Tymms, P., Vostanis, P., Belsky, J., Fonagy, P., Brown, A., Martin, A., Patalay, P. & Wolpert, W. (2013). The Development of a School-Based Measure of Child Mental Health. *Journal of Psychoeducational Assessment*, 31, 247-257.
- Department of Education. (2016). *Mental health and behaviour in schools: Departmental advice for school staff*. Retrieved from [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/508847/Mental\\_Health\\_and\\_Behaviour\\_-\\_advice\\_for\\_Schools\\_160316.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/508847/Mental_Health_and_Behaviour_-_advice_for_Schools_160316.pdf).
- Department of Health (DoH). (2011). *No health without mental health: A cross-government mental health outcomes strategy for people of all ages*. Retrieved from [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/213761/dh\\_124058.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/213761/dh_124058.pdf).
- DoH. (2015). *Future in Mind: Promoting, protecting and improving our children and young people's mental health and wellbeing*. Retrieved from:

<https://www.gov.uk/government/publications/improvingmental-health-services-for-young-people>.

DoH. (2017). *Transforming children and young people's mental health provision: A green paper*.

Retrieved from [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/664855/Transforming\\_children\\_and\\_young\\_people\\_s\\_mental\\_health\\_provision.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664855/Transforming_children_and_young_people_s_mental_health_provision.pdf).

Duncan, B. L., Sparks, J. A., Miller, S. D., Bohanske, R. T., & Claud, D. A. (2006). Giving youth a voice: A preliminary study of the reliability and validity of a brief outcome measure for children, adolescents and caretakers. *Journal of Brief Therapy*, 5, 71-88.

DuPaul, G. J. & Weyandt, L. L. (2006). School-based Intervention for Children with Attention Deficit Hyperactivity Disorder: Effects on academic, social, and behavioural functioning. *International Journal of Disability, Development and Education*, 53, 161-176.

Fazel, M., Hoagwood, K., Stephan, S. & Ford, T. (2014). Mental health interventions in schools in high-income countries. *The Lancet Psychiatry*, 1, 377 – 387.

Geng, G. (2011). Investigation of Teachers' Verbal and Non-verbal Strategies for Managing Attention Deficit Hyperactivity Disorder (ADHD) Students' Behaviours within a Classroom Environment. *Australian Journal of Teacher Education*, 36, 17-30.

Hassiotis, A., Serfaty, M., Azam, K., Martin, S., Strydom, A., & King, M. (2012). *A Manual of Cognitive Behaviour Therapy for People with Learning Disabilities and Common Mental Disorders: Therapist Version*. Retrieved from <file:///C:/Users/k1505463/Downloads/manual-cbt-id.pdf>.

Hirshfeld-Becker, D. R. & Biederman, J. (2002). Rationale and Principles for Early Intervention with Young Children at Risk for Anxiety Disorders. *Clinical Child & Family Psychology Review*. 5, 161-172.

Lenehan, C. & Geraghty, M. (2017). *Good intentions, good enough?: A review of the experiences and outcomes of children and young people in residential special schools and colleges*. Retrieved from [http://dera.ioe.ac.uk/30465/1/Good\\_intentions\\_good\\_enough\\_-\\_a\\_review\\_of\\_residential\\_special\\_schools\\_and\\_colleges.pdf](http://dera.ioe.ac.uk/30465/1/Good_intentions_good_enough_-_a_review_of_residential_special_schools_and_colleges.pdf).

- Maddox, L., Jolley, S., Laurens, K. R., Hirsch, C., Hodgins, S., Browning, S., Bravery, L., Bracegirdle, K., Smith, P. & Kuipers, E. (2013). Cognitive Behavioural Therapy for Unusual Experiences in Children: A Case Series. *Behavioural and Cognitive Psychotherapy*, 41, 344-358.
- McCrone, P., Dhanasiri, S., Patel, A., Knapp, M. & Lawton-Smith, S. (2008). *Paying the price: The cost of mental health care in England to 2026*. Retrieved from [https://www.kingsfund.org.uk/sites/default/files/Paying-the-Price-the-cost-of-mental-health-care-England-2026-McCrone-Dhanasiri-Patel-Knapp-Lawton-Smith-Kings-Fund-May-2008\\_0.pdf](https://www.kingsfund.org.uk/sites/default/files/Paying-the-Price-the-cost-of-mental-health-care-England-2026-McCrone-Dhanasiri-Patel-Knapp-Lawton-Smith-Kings-Fund-May-2008_0.pdf).
- Meltzer, H., Gatward, R., Corbin, T., Goodman, R. & Ford, T. (2003). *Persistence, Onset, Risk Factors and Outcomes of Childhood Mental Disorders*. Retrieved from [http://www.dawba.com/abstracts/B-CAMHS99+3\\_followup\\_report.pdf](http://www.dawba.com/abstracts/B-CAMHS99+3_followup_report.pdf).
- National CAMHS Support Service. (2011). *Better Mental Health Outcomes for Children and Young People: A resource directory for commissioners*. Retrieved from <https://www.rcpsych.ac.uk/pdf/Bettermentalhealthoutcomesforchildrenandyoungpeo.pdf>.
- National Institute for Health and Clinical Excellence (NICE). (2008). *Social and emotional wellbeing in Primary Education*. Retrieved from <https://www.nice.org.uk/guidance/ph12>.
- NICE. (2013). *Antisocial behaviour and conduct disorders in children and young people: recognition and management*. Retrieved from <https://www.nice.org.uk/guidance/cg158/chapter/1-Recommendations#identifying-effective-treatment-and-care-options>.
- Neil, A. L. & Christensen, H. (2009). Efficacy and effectiveness of school-based prevention and early intervention programs for anxiety. *Clinical Psychology Review*, 29, 208–15.
- Public Health England. (2018). *Prevention Concordat for Better Mental Health*. Retrieved from <https://www.gov.uk/government/publications/prevention-concordat-for-better-mental-health-consensus-statement/prevention-concordat-for-better-mental-health>.
- Rotheram-Fuller, E. & Macmullen, L. (2011). Cognitive-Behavioral Therapy for Children with Autism Spectrum Disorders. *Psychology in the Schools*, 48, 263-271.

South London and Maudsley NHS Trust. (2018). *Cues-Ed*. Retrieved April 23, 2018, from <http://www.cues-ed.co.uk/>.

Stallard, P., Skryabina, E., Taylor, G., Phillips, R., Daniels, H., Anderson, R. & Simpson, N. (2014). Classroom-based cognitive behaviour therapy (FRIENDS): A cluster randomised controlled trial to Prevent Anxiety in Children through Education in Schools (PACES). *The Lancet Psychiatry*, 1, 185-192.

Wells, J., Barlow, J. & Stewart-Brown, S. (2003). A systematic review of universal approaches to mental health promotion in schools. *Health Education*, 103, pp.197-220.

Werner-Seidler, A., Perry, Y., Caele, A. L., Newby, J. M. & Christensen, H. (2017). School-based depression and anxiety prevention programs for young people: A systematic review and meta-analysis. *Clinical Psychology Review*, 51, 30-47.

## 7. APPENDICES

### 7.1. Table of CBT adaptations extracted from the literature.

<b>ASD (Rotheram-Fuller &amp; MacMullen, 2011)</b>
<ul style="list-style-type: none"><li>- Use of visual aids.</li><li>- Use of hands-on activities.</li><li>- Written worksheets to introduce new concepts.</li><li>- Provide creative outlets for expression (e.g., photography, drawing).</li><li>- Focus on strengths and expanding on current areas of interest.</li><li>- Socratic questions that incorporate hints of the correct answer to provide opportunities to put concepts into their own words.</li><li>- Exposure and practice of new skills.</li><li>- Provide multiple opportunities for repetition and practice.</li><li>- Model-specific feedback.</li><li>- Incorporate social reinforcement.</li><li>- Video modelling and video activities.</li><li>- In vivo rehearsal.</li><li>- Group therapy for practice opportunities.</li><li>- Promoting generalization of skills to new contexts.</li><li>- Parent participation interwoven into sessions to improve generalization of learned skills.</li><li>- In vivo rehearsal to address challenges in the most relevant setting.</li><li>- On-site coaching before and after an interaction or behaviour to improve generalization to nonclinical settings.</li></ul>
<b>ADHD (DuPaul &amp; Weyandt, 2006; Geng, 2011)</b>
<ul style="list-style-type: none"><li>- Choice-making interventions allow students to choose from two or more concurrently presented classroom activities (e.g., menu of academic tasks).</li><li>- Modify task assignments: written assignments may be modified by reducing their overall length and/or breaking the assignment into smaller sub-units and providing students with a brief break after the completion of each sub-unit.</li><li>- Active teaching of classroom rules.</li><li>- Simply-worded rules/instructions, to remind when rules have been followed.</li><li>- Token reinforcement: children earn immediate reinforcers (e.g., stickers, points) for meeting behavioural expectations.</li><li>- Clear organisational routines and procedures.</li><li>- Making frequent use of students' names.</li><li>- Standing close when giving instructions and maintaining eye contact.</li><li>- Using specific and direct instructions.</li><li>- Illustrating and writing instructions in addition to speaking them.</li><li>- Develop rapport with the students and treating them with respect.</li><li>- Accompanying speech with more visual gestures, as it can provide additional information and appear less abstract compared to spoken instructions.</li><li>- Voice control (low to loud volume, firmness, tone and pace).</li><li>- Short phrases, repeated instructions.</li><li>- Gently touching the student or pointing out the important information.</li></ul>

**Learning Difficulties (Hassiotis et al., 2012)**

- More specific and didactic teaching.
- Presenting key concepts in extremely concrete ways.
- Extra support in the form of visual aids such as pictures, drawings, and signs for certain tasks.
- Taking therapy at a slower pace.
- Using repetition.
- Encouraging “overlearning” in some scenarios.
- Simple language.
- Allowing time to respond.
- Speak slowly.
- Short, simple questions.
- Link explanation to everyday things/examples.
- Write down key info and share with key support.
- Focus on non-verbal – pics and communication.

**Other SEMH adoptions (Council for the Curriculum, Examinations and Assessment [CCEA], 2016)**

- Modelling respectful behaviour.
- Praise is offered frequently.
- Incorporating activities that allow success.
- Use of simple and clear language.
- Providing explicit instructions that do not rely on assumptions.
- Not referencing “home” or “parents” as not all children will live here/with these people.
- Be inclusive.
- Provide attention and warmth.
- Incorporate play and interactive methodologies.
- Include opportunities to apply learnt skills to other contexts.
- Create a high-level of engagement from students, teachers, and carers.

## SERVICE-RELATED PROJECT

---

# **Promoting Recovery within Croydon IAPT:**

Which clients do not recover?

---

**Supervised by:** Dr Jane Ellis



## ABSTRACT

---

Improving Access to Psychological Therapies (IAPT) was a government initiative to increase the availability of psychological therapies within the NHS for people with common mental health problems. An integral part of the IAPT programme is detailed outcome monitoring, which is used to evaluate service performance by using these outcome scores to determine client recovery status (i.e. clients are deemed as recovered once their outcome scores fall below the “caseness” threshold). However, reporting on these recovery rates does not provide information on why clients may not be recovering from IAPT services, whereby this knowledge could help enhance greater service rates of recovery. Therefore, this was a primary aim of this service evaluation, and a secondary aim was to identify how useful the current recovery rate calculation was at measuring service performance. This evaluation studied all individuals who were discharged from the Croydon IAPT service after completing treatment between 1<sup>st</sup> January and 31<sup>st</sup> December 2015 ( $N = 566$ ), and retrospectively collected routine data from the electronic patient system ('IAPTus') on several demographic, outcome, and service-related variables. The findings demonstrated that at this service, several variables differed significantly between those who recovered and those who did not (including: language; long-term condition; employment; presenting problem; risk; and the number of appointments). The unrecovered group was also found to have significantly higher outcome scores on triage and discharge compared to the recovered group, as well as demonstrating significantly less change in scores over treatment. Moreover, treatment type, referral, employment, and low GAD-7 scores at triage were each found to predict recovery. Finally, the “moving to recovery” calculation yielded the highest recovery rate figure, compared to the “reliable change index” and “reliable recovery” methods, although the utility of this measure is limited. The limits and implications of the findings are discussed, and recommendations for promoting recovery at Croydon IAPT are suggested.

## CONTENTS

---

<b>1. INTRODUCTION .....</b>	<b>117</b>
1.1. The IAPT Programme.....	117
1.2. Outcome Measures and Recovery .....	117
1.3. Current Predictors of Recovery .....	119
1.4. The Current Evaluation.....	120
<b>2. METHODS .....</b>	<b>121</b>
2.1. Sample.....	121
2.2. Measures.....	122
2.2.1. The Patient Health Questionnaire – 9 (PHQ-9): .....	123
2.2.2. The Generalised Anxiety Disorder - 7 (GAD-7):.....	123
2.3. Procedure .....	124
2.4. Ethics .....	124
2.5. Service User Involvement.....	124
2.6. Statistical Analysis .....	124
<b>3. FINDINGS .....</b>	<b>126</b>
3.1. Demographic Variables .....	126
3.1.1. Age.....	126
3.1.2. Gender .....	127
3.1.3. Ethnicity.....	128
3.1.4. Sexuality.....	128
3.1.5. Language.....	130
3.1.6. Disability .....	131
3.1.7. Long-Term Conditions (LTC) .....	132
3.1.8. Employment Status on Discharge.....	133
3.1.9. Medication Use on Discharge.....	134
3.2. Service Contact Variables .....	135
3.2.1. Referral Source .....	135
3.2.2. Number of Previous Referrals .....	136
3.2.3. Primary Presenting Problem.....	137
3.2.4. Risk Level .....	139
3.2.5. Time on Waiting List .....	140
3.2.6. Therapy Type .....	141
3.2.7. Group Therapy Attendance .....	142
3.2.8. Therapeutic Contact Type .....	142
3.3. Outcome Measure Variables .....	144

3.4. Recovery Rate .....	146
3.4.1. Moving to Recovery .....	146
3.4.2. Reliable Change Index .....	146
3.4.3. Reliable Recovery Index .....	146
<b>4. DISCUSSION .....</b>	<b>147</b>
4.1. Summary .....	147
4.2. Limitations.....	149
4.3. Implications and Recommendations.....	150
4.4. Conclusions .....	150
<b>5. DISSEMINATION .....</b>	<b>154</b>
<b>6. LEADERSHIP .....</b>	<b>155</b>
<b>7. REFERENCES.....</b>	<b>156</b>
<b>8. APPENDICES .....</b>	<b>159</b>
8.1. PHQ-9 and GAD-7 Questionnaires .....	159
8.2. 'Key Findings' hand-out.....	160
8.3. PowerPoint presentation for service staff .....	163

## 1. INTRODUCTION

---

### 1.1. The IAPT Programme

In 2007 and on “World Mental Health Day”, the UK labour Government announced a new initiative named Improving Access to Psychological Therapies (IAPT) for the treatment of people with common mental health problems (i.e. anxiety and depression disorders). It was created to significantly increase the availability of evidence-based psychological therapies within the National Health Service (NHS), namely Cognitive Behavioural Therapy (CBT), which is the recommended treatment by the National Institute of Health and Clinical Excellence (NICE) (Department of Health [DoH], 2012). The main influencing factors which led to the development of such a large-scale initiative came from both economists and clinical researchers, who demonstrated that IAPT would have several economic and social benefits for the nation. For example, by increasing access to the programme it would reduce suffering and pay for itself through the reduction of depression/anxiety-related public costs (welfare benefits and medical costs) and increasing revenues (taxes from return to work, increased productivity etc.) (Gyani, Shafran, Layard & Clark, 2013).

In 2011, and following the success of a pilot study, the plans for the IAPT initiative to become a reality were published in the DoH’s document “Talking Therapies: A Four-Year Plan of Action” (2011<sub>a</sub>). This plan was an accompaniment to the Government’s mental health strategy “No Health, Without Mental Health”, which aims to improve outcomes for people with mental health problems through high-quality services that are equally accessible to all (DoH, 2011<sub>b</sub>). The plan stipulates that by the end of the fourth year: (1) the nationwide roll-out of the programme would be completed; (2) initiations for the programme to be expanded to children and young people would have begun; and (3) models of care for people with long-term physical conditions, medically unexplained symptoms, or severe mental illness would be developed (IAPT Programme, 2016).

The IAPT Programme (2016) currently report that IAPT is now established in every area of England with an estimated 6,000 therapists trained.

### 1.2. Outcome Measures and Recovery

An integral part of the IAPT programme is the detailed outcome monitoring and continued evaluation of the service (Gyani et al., 2013). This is namely achieved using session-by-session outcome measures, which are then shared with clients during sessions to demonstrate their

clinical progress (IAPT National Programme Team, 2011). Moreover, a client's score is also used to inform a judgement of whether a client is above or below clinical "caseness" (i.e. their symptoms are sufficiently severe to be considered a clinical problem), where an individual is deemed as "recovered" once they have moved below this caseness threshold (Griffiths & Steen, 2013; Clark & Oates, 2014). Therefore, these outcomes are also used by managers and service commissioners to assess service performance, and to demonstrate the direct return on the investment made in services (IAPT National Programme Team, 2011).

Typically, IAPT use the "Moving to Recovery" indicator to calculate their recovery rate. This is the proportion of clients who complete treatment and move from being at clinical caseness at the beginning of therapy, to below the caseness threshold at the end of therapy (Griffiths & Steen, 2013). However, this method ignores those clients who do not end up falling below the threshold at the end of treatment, but are still able to achieve worthwhile benefits, such as being better able to recognise and manage their illness from techniques learnt in therapy (DoH, 2012). So, by using the "Moving to Recovery" indicator, it means that those who show significant improvements in scores but remain over the caseness threshold are not counted in the final figure, but those who make small changes across the threshold (which can be achieved without intervention) are (Clark & Oates, 2014).

Alternatively, Gyani, Shafran, Layard, and Clark (2011) outline a recovery calculation method which considers whether the change in scores from pre- to post-treatment exceeds the measurement error of the relevant scale, and hence can be considered statistically reliable. This is called the "Reliable Change Index", and by adopting this method it means that only "real change" is acknowledged and thus illustrates the number of people who show any degree of real benefit while being treated in an IAPT service (Clark & Oates, 2014). If clients meet the criteria for both the "Moving to Recovery" and "Reliable Change Index" measures, this is considered as "Reliable Recovery" (Health and Social Care Information Centre, 2015).

IAPT recovery rates have typically been found to vary between 41-56% (Chan & Adams, 2014; Clark et al., 2009; DoH, 2012; Gyani et al., 2013; Radhakrishnan et al. 2013; Richards & Borglin, 2011), where the majority fall below the IAPT target of 50% (National IAPT Programme Team, 2012). However, if recovery were defined by the reliable change index, then some of these figures would increase to over 60% (i.e. achieving significant improvements in symptoms but not achieving the technical definition of recovery) (DoH, 2012; Gyani et al., 2013). The discrepancies between rates are still observed in the most recent IAPT figures (2014/15), which

demonstrate a recovery rate of 44.8%, a reliable recovery of 42.8%, and reliable change of 60.8% (Health and Social Care Information Centre, 2015).

Griffiths and Griffiths (2015) further raise questions about the implications for clients scoring at the top end of outcome measures, as they are required to show an even greater amount of improvement on measures to achieve “recovery”. Their research demonstrated that only a third of clients scoring “severe” on the standard IAPT outcome measures recovered when using the traditional recovery rate. Consequently, as this distinct group are less likely to recover, Griffiths and Griffiths query what IAPT services will do to provide more effective treatment for them. Atkinson (2015) argues that none of the recovery estimates actually present the true recovery figure in the context of the high number of referrals IAPT receives. This figure would be significantly lower than the typical figure of above 40%, and the difference between the method favoured by the IAPT programme and the proportion of all referrals is too large to be ignored – especially from the view of commissioners and GPs making the referrals (Griffiths and Steen, 2013).

### 1.3. Current Predictors of Recovery

The identification of factors likely to predict recovery can be extremely valuable to any service, as it can provide guidance on service developments and inclusion/exclusion criteria, as well as improving patient experience. Since the development of IAPT, there have been several studies investigating such predictors, including Gyani and colleagues (2011) who developed a comprehensive report analysing the data from the first year of IAPT services from 32 sites. They aimed to identify the factors which may explain the variability in recovery. Their findings suggested that recovery was more likely in patients who: had lower symptom severity at the start of treatment; had a diagnosis of a depressive episode, mixed anxiety and depressive disorder, generalised anxiety disorder, or post-traumatic stress disorder; and received a greater number of therapy sessions. The likelihood of recovery was also explored on a service level, whereby services with better recovery figures were found to have treated more patients, had a greater proportion of sessions delivered by therapists banded at level 7 or above (‘Agenda for Change’ pay scale), and stepped up a greater proportion of clients.

Vaillancourt, Manley, and McNulty (2015) also reviewed factors associated with recovery within one London IAPT service, by comparing data from two-time points of when the service obtained a high versus low recovery rate. They measured patient symptom severity, waiting time, and length of treatment, although none of these factors were found to explain the reduction in

recovery rate across the two-time periods. Nonetheless, they did find that patients with higher symptom severity and those who received a smaller length of treatment were less likely to recover.

Further, Saunders, Cape, Fearon, and Pilling (2016) were able to identify several patient profiles which predicted recovery within services delivering psychological treatment for anxiety and depression. Patients classified in the “LP1” group were the most likely to recover and respond to treatment. They presented to the service with: low symptom severity, a relatively high level of functioning, fewer phobic symptoms, and were less likely to receive public welfare benefits and be prescribed medication compared to the overall sample. The worst outcomes were found in the “LP7” group, where recovery was 4.9 times more likely in LP1 compared to LP7. LP7 patients had the highest intake symptom severity, a high probability of receiving public welfare benefits, prescribed medication, and had phobia symptoms.

#### 1.4. The Current Evaluation

Despite the recovery rate being an important outcome to report to senior management and commissioners, it fails to identify the reasons *why* clients may not be recovering from the Croydon IAPT service, and it is not understood how reflective the current measure of recovery is. If the Croydon service could develop a better understanding of the factors that are or are not associated with recovery, it would be able to educate its therapists and be better able to identify and manage these within treatment – hopefully having a positive impact on the service’s recovery performance and patient experience. By understanding how useful the current measure of recovery is, it would allow the Croydon service to consider other additional methods when assessing actual recovery performance.

Taken together, the primary aim of this service evaluation was to identify the factors associated with recovery in Croydon IAPT. A secondary aim was to consider whether the “Moving to Recovery” calculation is a useful measure of recovery for this service.

## 2. METHODS

This service evaluation project used a comparative cross-sectional design to understand the differences between those clients who did and did not recover from the Croydon IAPT service within the year of 2015. A client's recovery outcome ("Recovered" or "Unrecovered") was determined by their score on routinely used outcome measures at discharge, regardless to what their scores were on triage.

"Recovered" clients had discharge scores below the caseness threshold on both the PHQ-9 and GAD-7, whereas "Unrecovered" clients had at least one of their discharge PHQ-9/GAD-7 scores within the caseness threshold (Table 9). Please see the "Measures" section for more information on the PHQ-9/GAD-7.

Overall, 64.8% ( $N = 367$ ) of the total sample were considered as "Recovered", leaving 35.2% ( $N = 199$ ) "Unrecovered".

**Table 9: The range and caseness threshold for the PHQ-9 and GAD-7 (Clark and Oates, 2014).**

	Range	Caseness Threshold
<i>PHQ-9</i>	0-27	10 and above
<i>GAD-7</i>	0-21	8 and above

### 2.1. Sample

The sample consisted of any individual who was discharged from the Croydon IAPT service after completing treatment between 1<sup>st</sup> January and 31<sup>st</sup> December 2015. Six clients were excluded from the sample due to not actually receiving any treatment from the service, such as those who only had scores from their triage contact and/or were referred on to other services. Two clients also had two care episodes during 2015 so these were added to the sample. Subsequently this produced a total sample of 566 clients.

In terms of the sample's demographics, 64.7% ( $N = 366$ ) of the sample were female and 35.3% ( $N = 200$ ) male. The mean average age of clients was 43.2 years ( $SD = 14.9$ ) with a range of 17-84 years. Over half of the sample identified their ethnicity as 'White British' (55.1%,  $N = 312$ ),



and the following most prevalent ethnicities were 'Black or Black British' (12.0%,  $N = 68$ ), 'Asian or Asian British' (9.4%,  $N = 53$ ), and 'White Other' (5.7%,  $N = 32$ ). However, for 6% ( $N = 34$ ) of clients their ethnicity was not reported on, 1.9% ( $N = 11$ ) chose not to state this, and 0.9% ( $N = 5$ ) did not know. The most common primary presenting problem was depression/low mood (44.3%,  $N = 251$ ), followed by anxiety (23.3%,  $N = 132$ ), and panic (6.4%,  $N = 36$ ).

## 2.2. Measures

Clients were compared on variables across three main areas, that is: 1) Demographic Information, 2) Service Contact, and 3) Outcome Measures. Please see Table 10 below for the specific variables collected in each area per client.

**Table 10: A list of the categorised data variables collected per client.**

Demographic Information	Service Contact	Outcome Measures
– Gender	– Referral Type	– PHQ-9 score on Triage
– Age on discharge	– Number of Previous Referrals	– PHQ-9 score on Discharge
– Ethnicity	– Primary Presenting Problem	– Difference in PHQ-9 score from triage to discharge
– Sexuality	– Risk Level at Triage	– GAD-7 score on Triage
– Language	– Time spent on the Waiting List	– GAD-7 score on Discharge
– Disability	– Type of Individual Therapy Received	– Difference in GAD-7 score from Triage to Discharge
– Long-Term Condition	– Attendance to a Group	
– Employment on Discharge	– Number of attended Individual therapeutic contacts	
– Medication use on Discharge	– Number of times the client 'Did Not Attend' (DNA)	
	– Number of times the client was late	
	– Number of times the client cancelled	
	– Number of times the service cancelled	
	– Total number of booked appointments	

#### 2.2.1. The Patient Health Questionnaire – 9 (PHQ-9):

The PHQ-9 (Kroenke, Spitzer, & Williams, 2001) is a brief self-report questionnaire consisting of nine statements which assess the severity of depressive symptoms (Appendix 1). Each statement represents a typical symptom of depression. Clients are required to rate each of the nine statements according to how much they occurred in the last two weeks. This scale ranges from 0-3 representing responses of “not at all” to “nearly every day”. The authors of the measure suggest that higher scores indicate more severe depressive symptoms and developed the following score classifications:

- 0-4 = Healthy
- 5-9 = Mild
- 10-14 = Moderate
- 14-19 = Moderately Severe
- 20 or above = Severe

It has also been found to be a reliable and valid measure of depression (Kroenke et al., 2001).

#### 2.2.2. The Generalised Anxiety Disorder - 7 (GAD-7):

The GAD-7 (Spitzer, Kroenke, Williams, & Löwe, 2006) is another brief self-report questionnaire like the PHQ-9. However, it consists of seven statements and assesses the severity of anxiety symptoms (Appendix 1). Each statement represents a typical symptom of anxiety, whereby clients rate the frequency of each symptom in the last two weeks according to the same 0-3 likert scale as the PHQ-9. Similarly, the authors suggest that higher scores indicate more severe symptoms on anxiety, and provide the following classifications of scores:

- 0-4 = Healthy
- 5-9 = Mild
- 10-14 = Moderate
- 15 or above = Severe

The GAD-7 is also considered a valid and efficient tool for screening for GAD and assessing its severity in clinical practice and research (Spitzer et al., 2006).

### 2.3. Procedure

All data was routinely collected prior to this service evaluation and stored within the service's electronic patient record system "IAPTus". The researcher accessed each client's record to extract the relevant data and recorded it in a password protected Microsoft Excel Spreadsheet. This was later transferred without patient identifiable information (i.e. client identification numbers) to a SPSS (Statistical Package for the Social Sciences) database for further statistical analysis.

### 2.4. Ethics

Information governance and project approval was obtained from the Clinical Governance Team within the Mood, Anxiety, and Personality Clinical Academic Group (MAP CAG) of the South London and Maudsley NHS Foundation Trust. Individual consent from each client was not obtained or required. All data was stored anonymously to ensure confidentiality.

### 2.5. Service User Involvement

Service Users and Carer's within the MAP CAG were consulted about the service evaluation at two of their monthly advisory meetings. The first meeting was attended within the project's development phase (prior to receiving project approval) with the intention to obtain their views on how best to achieve the aims of the project and whether, in their view, any key data variables associated with recovery were missing. No changes to the proposal were agreed. The second meeting was attended after the findings had been analysed. This was to feedback the results and gather any recommendations/actions that they believed the service should consider in light of such findings. Suggestions were then fed back to the service during a team meeting.

### 2.6. Statistical Analysis

All analysis was conducted using SPSS 22.0, a statistical software package for Windows. All data was analysed using descriptive statistics - that is to say, using measures of central tendency and dispersion (usually the mean average (*M*) and standard deviation (*SD*) respectively), as well as frequencies to identify any patterns within data.

Subsequent inferential statistical analysis was also carried out to understand whether any of the patterns identified were *statistically relevant* and could be generalised to the whole population. These tests varied according to the data being analysed and desired analytical

outcome. For example, independent T-Tests were used to identify whether any numerical data differed between the two recovery groups (e.g. outcome scores, number of sessions, time on waiting list etc.); repeated-measures T-Tests were used to identify whether any numerical data from the same client differed between two time points (e.g. their outcome scores on triage and discharge); Chi-Square tests were used to identify whether any categorical data differed between the two recovery groups (e.g. ethnicity, employment, referral source etc.); and a multivariate binary logistic regression was used to understand whether each variable could predict whether someone recovered from Croydon IAPT or not.

### 3. FINDINGS

---

Below are several headings which correspond to each variable that was measured. Here, any data patterns will be described, which will be followed by comments about whether any statistically-relevant differences exist between the two recovery groups, and whether this variable was found to statistically predict recovery group. Overall, the binary logistic regression model was statistically significant, indicating that the predictors as a set reliably distinguished between those who did and did not recover ( $\chi^2(29) = 75.36, p < 0.01$ ). The model explained 26.0% of the variance in recovery (Nagelkerke's  $R^2$ ) and correctly classified 75.1% of cases (47.2% for the unrecovered group and 89.5% for the recovered group).

#### 3.1. Demographic Variables

##### 3.1.1. Age

The mean average age of clients in the unrecovered group was 43.3 years, which is comparable to recovered group, whose mean age was 43.2 (Table 11). The Standard Deviation ( $SD$ ) and range were also comparable.

**Table 11: The mean age in years, standard deviations, and ranges for each recovery group.**

	<i>M</i>	<i>SD</i>	Range
<i>Unrecovered</i>	43.3	14.3	18-84
<i>Recovered</i>	43.2	15.2	17-83

##### **Are these differences significant?**

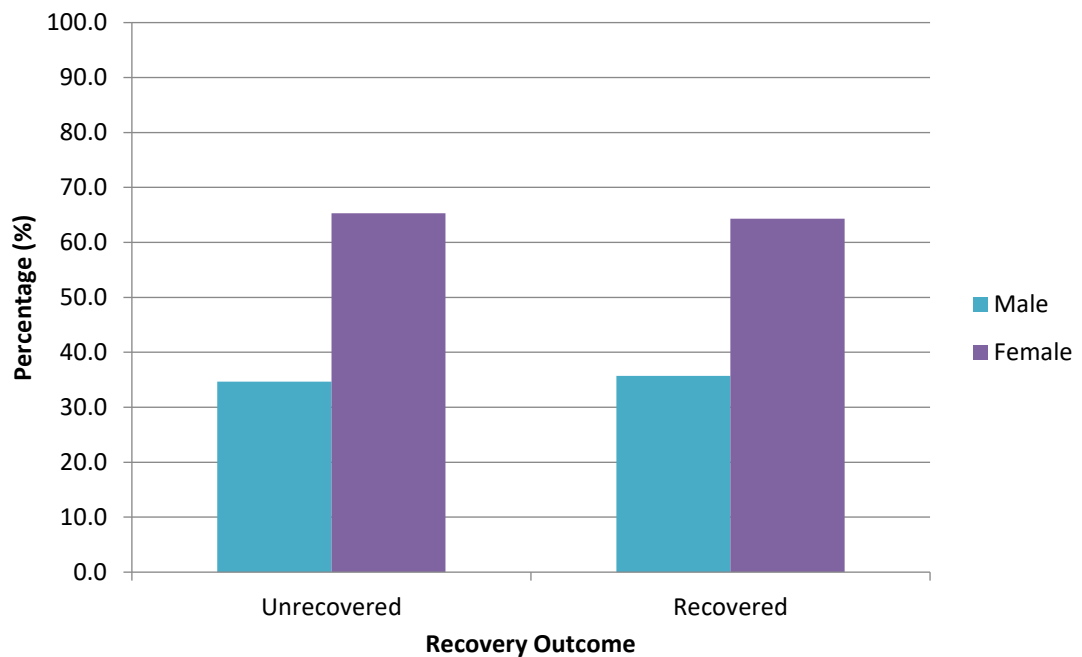
An independent T-Test was conducted to examine whether the difference was statistically significant. There was no significant difference found in age between those who recovered and those who did not ( $t(564) = 0.13, p = 0.90$ ).

##### **Does age predict whether someone recovers or not?**

Within the binary logistic regression model, age was not a significant predictor of recovery ( $b = -0.01, Wald \chi^2(1) = 0.29, p = 0.59$ ).

### 3.1.2. Gender

The proportion of males and females in each of the recovery groups appeared to be reasonably similar, with only a 1.0% difference found between them. In the recovered group, 64.3% ( $N = 236$ ) were female and 35.7% ( $N = 131$ ) were male (Figure 12). In the unrecovered group, 65.3% ( $N = 130$ ) were female and 34.7% ( $N = 69$ ) were male.



**Figure 12: Bar graph to show the percentage (%) of clients who were male/female according to their recovery outcome (recovered/unrecovered).**

#### ***Are these differences significant?***

A chi-square test of independence was performed to examine the relationship between recovery and gender. The relationship between these variables was not significant ( $\chi^2(1) = 0.06$ ,  $p = 0.81$ ).

#### ***Does gender predict whether someone recovers or not?***

Within the binary logistic regression model, gender was not a significant predictor of recovery ( $b = 0.28$ ,  $\text{Wald } \chi^2(1) = 0.93$ ,  $p = 0.34$ ).

### 3.1.3. Ethnicity

Figure 13 below presents the proportions of ethnicities present in each recovery group. For both recovery groups, the ethnicity identified most frequently was 'White British' (57.8%,  $N = 212$  for the recovered group; 50.3%,  $N = 100$  for the unrecovered group), however there was 7.5% more 'White British' individuals in the recovered group. Moreover, the second most prevalent ethnicity within the recovered group was 'Black or Black British' (11.4%,  $N = 42$ ) followed by 'Asian or Asian British' (6.8%,  $N = 25$ ). This was the opposite for the unrecovered group, where the second most prevalent ethnicity was 'Asian or Asian British' (14.1%,  $N = 28$ ), followed by 'Black or Black British' (13.1%,  $N = 26$ ). Ethnicity that was not reported on equal between groups at 6.0%.

#### ***Are these differences significant?***

A chi-square test of independence was performed to examine the relationship between recovery and ethnicity. The relationship between these variables was not significant ( $\chi^2(1) = 3.17, p = 0.08$ ).

#### ***Does ethnicity predict whether someone recovers or not?***

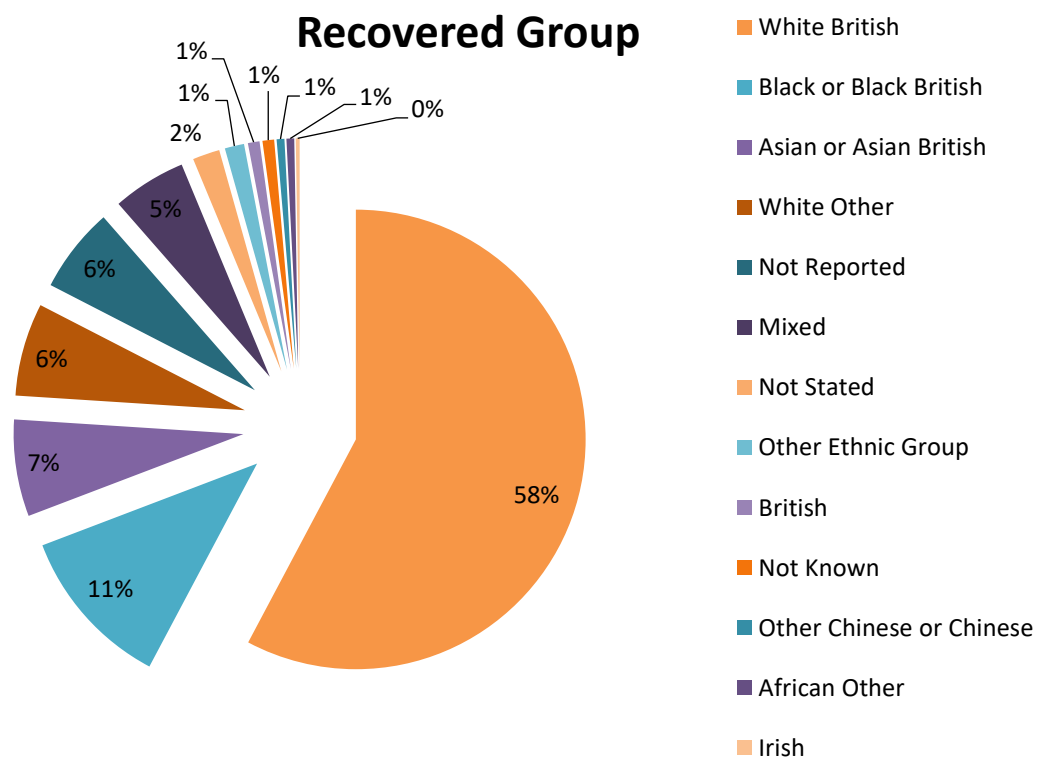
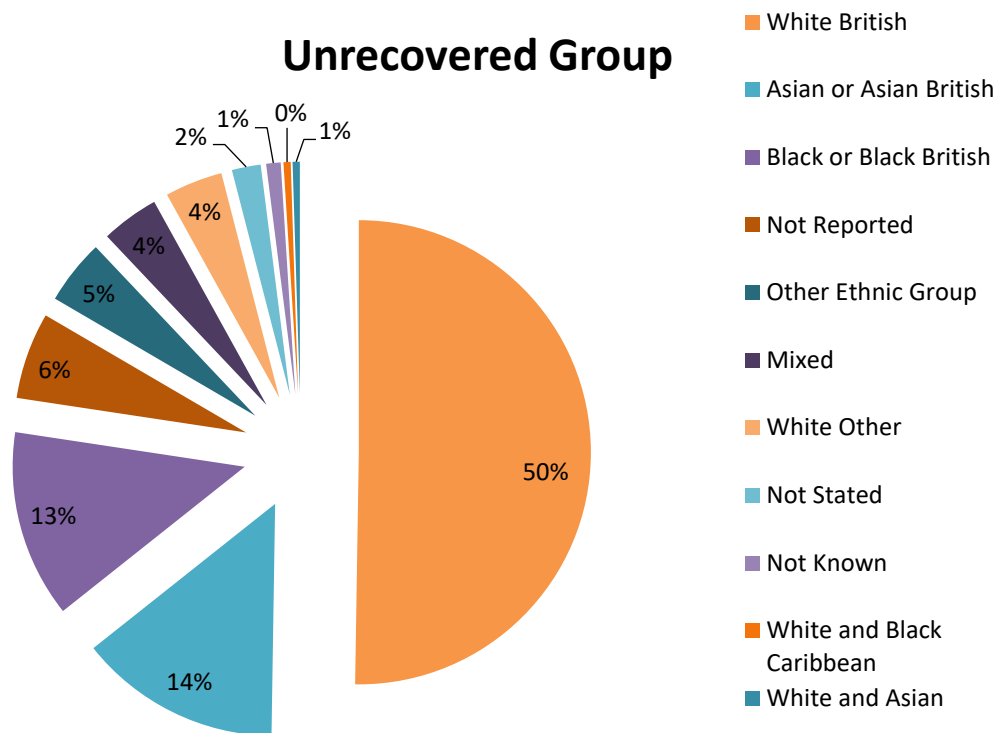
Within the binary logistic regression model, ethnicity was not a significant predictor of recovery ( $b = 0.08, \text{Wald } \chi^2(1) = 0.09, p = 0.77$ ).

### 3.1.4. Sexuality

As seen below in Figure 14, 'Heterosexual' was the most commonly described sexual orientation in both recovery groups (64.0%,  $N = 235$  in recovered group; 61.8%,  $N = 123$  in unrecovered group). Over a quarter of clients in both groups did not have this information reported on, but more so in the unrecovered group (25.9%,  $N = 95$  in recovered group; 30.2%,  $N = 60$  in unrecovered group).

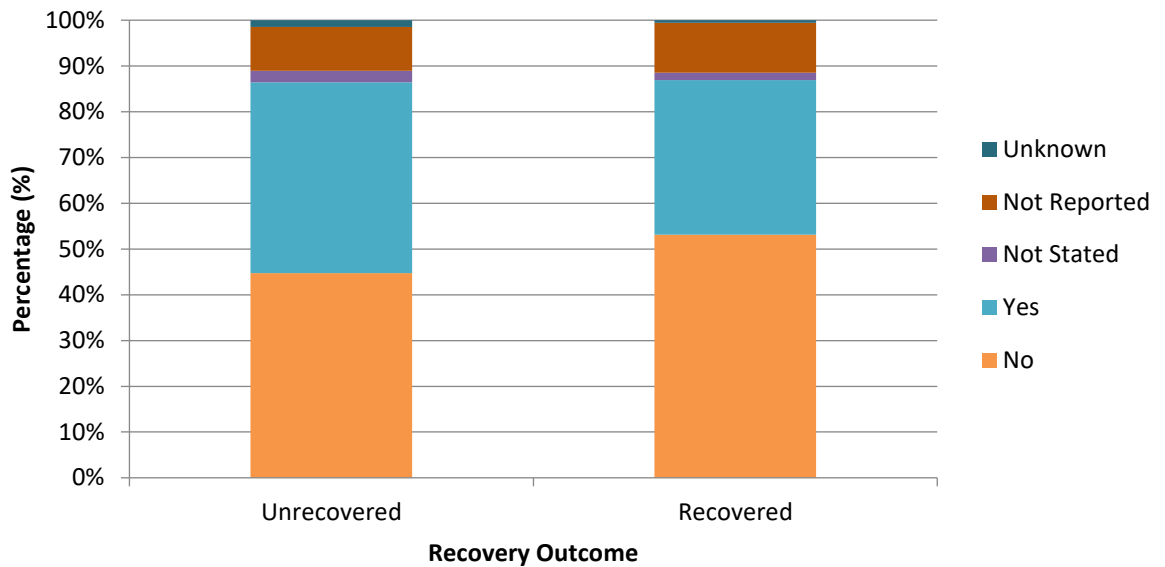
#### ***Are these differences significant?***

A chi-square test of independence was performed to examine the relationship between recovery and sexuality. The relationship between these variables was not significant ( $\chi^2(1) = 0.24, p = 0.60$ ).



**Figure 13: Pie graphs to show the proportions (%) of each ethnicity description in each recovery group.**





**Figure 14: Bar graph to show the spread of sexuality descriptions across recovery groups.**

#### ***Does sexuality predict whether someone recovers or not?***

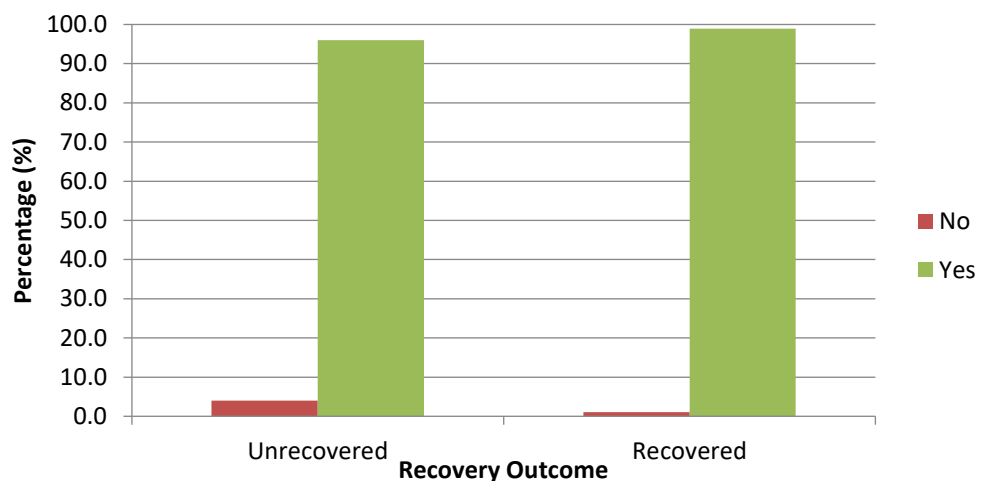
Within the binary logistic regression model, sexuality was not a significant predictor of recovery ( $b = 0.21$ ,  $\text{Wald } \chi^2(1) = 0.51$ ,  $p = 0.48$ ).

#### **3.1.5. Language**

In both recovery groups, over 95% of clients spoke English (Figure 15). This figure was greater in the recovered group (98.9%,  $N = 363$ ) compared to the unrecovered group (96.0%,  $N = 191$ ). There were double the number of clients who could not speak English in the unrecovered group ( $N = 8$ , 4.0%) compared to the recovered group ( $N = 4$ , 1.1%).

#### ***Are these differences significant?***

The relationship between recovery and language was found to be statistically significant ( $p = 0.03$ , Fisher's Exact Test). This suggested that language does have an effect of the frequency of recovery, whereby if you do not speak English you are much less likely to recover compared to those who do.



**Figure 15: Bar graph to present the percentage (%) of clients in each recovery group who could/could not speak English.**

***Does language predict whether someone recovers or not?***

Within the binary logistic regression model, language was not a significant predictor of recovery ( $b = 1.217$ ,  $Wald \chi^2(1) = 1.79$ ,  $p = 0.18$ ).

**3.1.6. Disability**

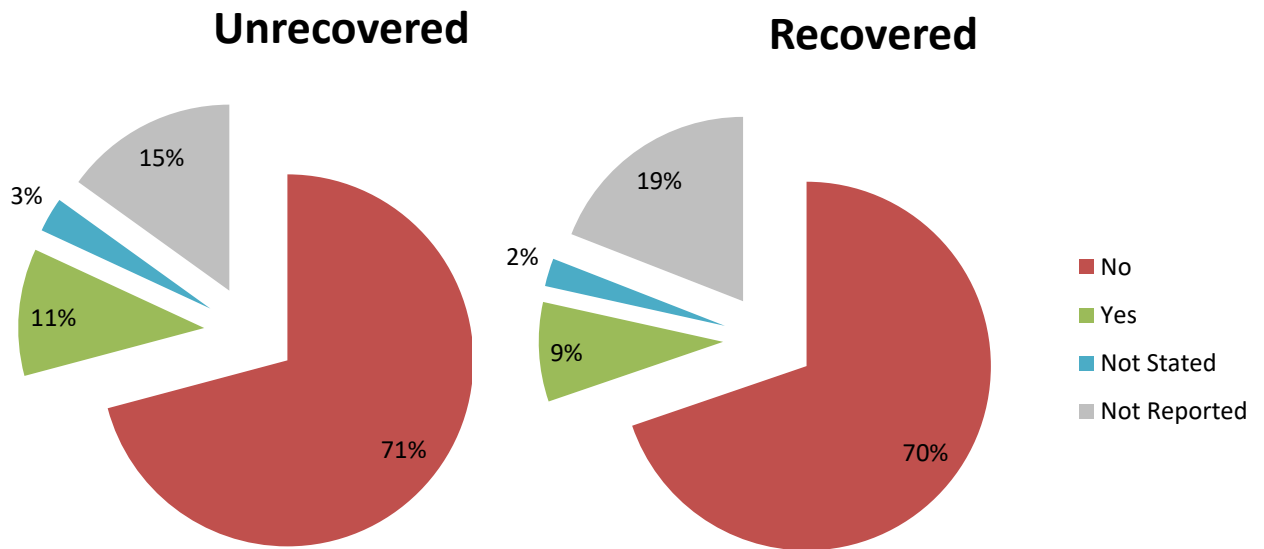
Approximately 70% of the sample in each recovery group reported having no perceived disability (70.9%,  $N = 141$  in the recovered group; 69.8%,  $N = 256$  in the unrecovered group). There was a higher reported prevalence of having a disability in the unrecovered group (11.1%,  $N = 22$ ) compared to the recovered group (8.7%,  $N = 32$ ). The remaining data was not reported on or clients did not wish to state (Figure 16).

***Are these differences significant?***

A chi-square test of independence was performed to examine the relationship between recovery and disability. The relationship between these variables was not significant ( $\chi^2(1) = 0.56$ ,  $p = 0.45$ ).

***Does having a disability predict whether someone recovers or not?***

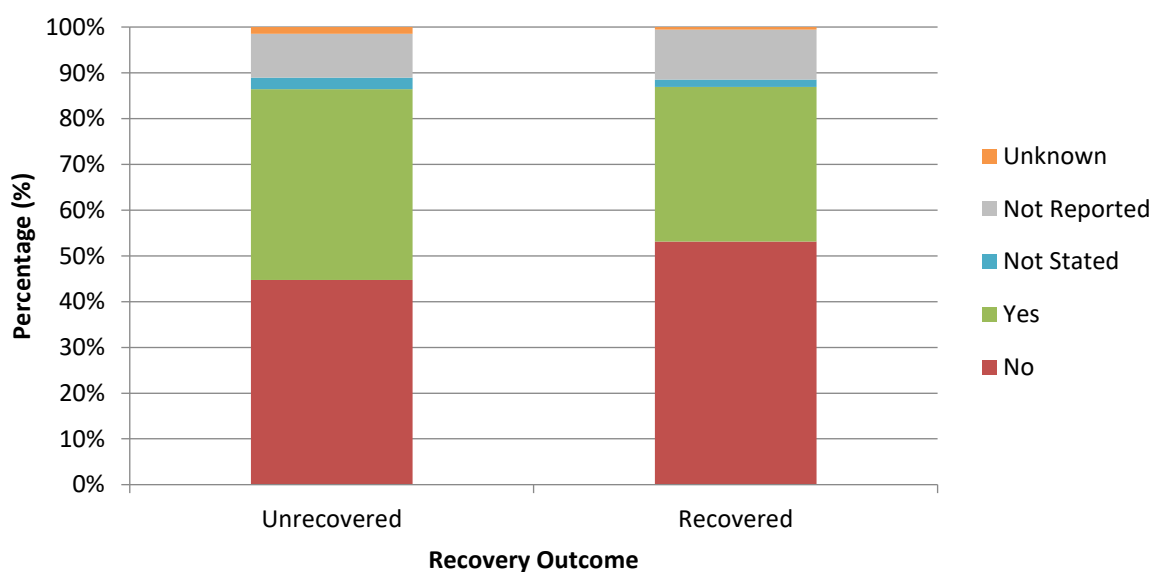
Within the binary logistic regression model, disability was not a significant predictor of recovery ( $b = -0.15$ ,  $Wald \chi^2(1) = 0.10$ ,  $p = 0.75$ ).



**Figure 16: Pie graphs to show the proportion (%) of clients who have a disability for each recovery group,**

### 3.1.7. Long-Term Conditions (LTC)

A greater proportion of clients with LTC were found in the unrecovered group (41.7%,  $N = 83$ ) compared to the recovered group (33.8%,  $N = 124$ ). Within the unrecovered group, the difference between those who do identify as having a LTC and those who do not is only 3% (“No” response - 44.7%,  $N = 89$ ) (Figure 17).



**Figure 17: Bar graph to show the proportion (%) of clients who consider themselves to have a LTC across recovery groups.**

### ***Are these differences significant?***

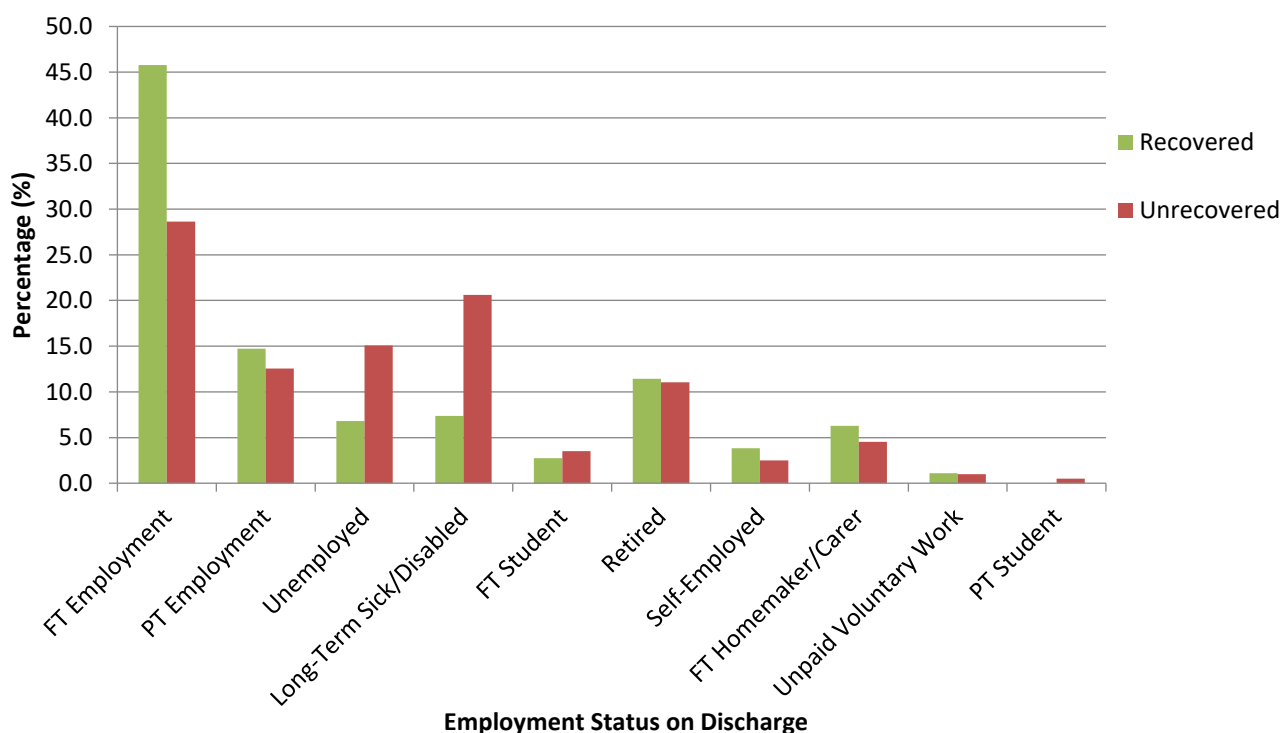
A chi-square test of independence was performed to examine the relationship between recovery and LTC. The relationship between these variables was found to be statistically significant ( $\chi^2(1) = 4.04, p = 0.05$ ), suggesting that significantly more clients recover when they do not have a LTC compared to when they do.

### ***Does having a LTC predict whether someone recovers or not?***

Within the binary logistic regression model, LTC was not a significant predictor of recovery ( $b = 0.12, \text{Wald } \chi^2(1) = 0.15, p = 0.70$ ).

#### 3.1.8. Employment Status on Discharge

For both recovery groups, having full-time employment was the most prevalent type of employment status, however this percentage was much less in the unrecovered group (45.8%,  $N = 168$  for the recovered group, 28.6%,  $N = 57$  for the unrecovered group). For unrecovered clients there was 8.3% more unemployment and 13.2% more clients who identified as 'Long-Term Sick/Disabled' compared to the recovered group. There was no more than a 2% difference between the recovery groups on all other employment statuses. Please see Figure 18.



**Figure 18: Bar graph to show the distribution (%) of employment status within each recovery group.**

### ***Are these differences significant?***

A chi-square test of independence was performed to examine the relationship between recovery and employment. The relationship between these variables was statistically significant ( $\chi^2(3) = 36.23, p < 0.01$ ). There were significantly fewer clients who were employed, and significantly more clients who identified as being unemployed or long-term sick/disabled within the unrecovered group than would be expected. In the recovered group, there were significantly less people who were long-term sick/disabled than what would be expected.

### ***Does Employment on discharge predict whether someone recovers or not?***

Within the binary logistic regression model, overall employment on discharge was found to statistically predict recovery ( $\text{Wald } \chi^2(3) = 10.29, p = 0.02$ ). More specifically, clients who were unemployed were 0.36 times less likely to recover ( $p = 0.02, 95\% \text{ CI} = 0.16 - 0.84$ ), and those who identified as being long-term sick/disabled were 0.34 times less likely to recover ( $p = 0.01, 95\% \text{ CI} = 0.15 - 0.77$ ) compared to those who were employed.

#### 3.1.9. Medication Use on Discharge

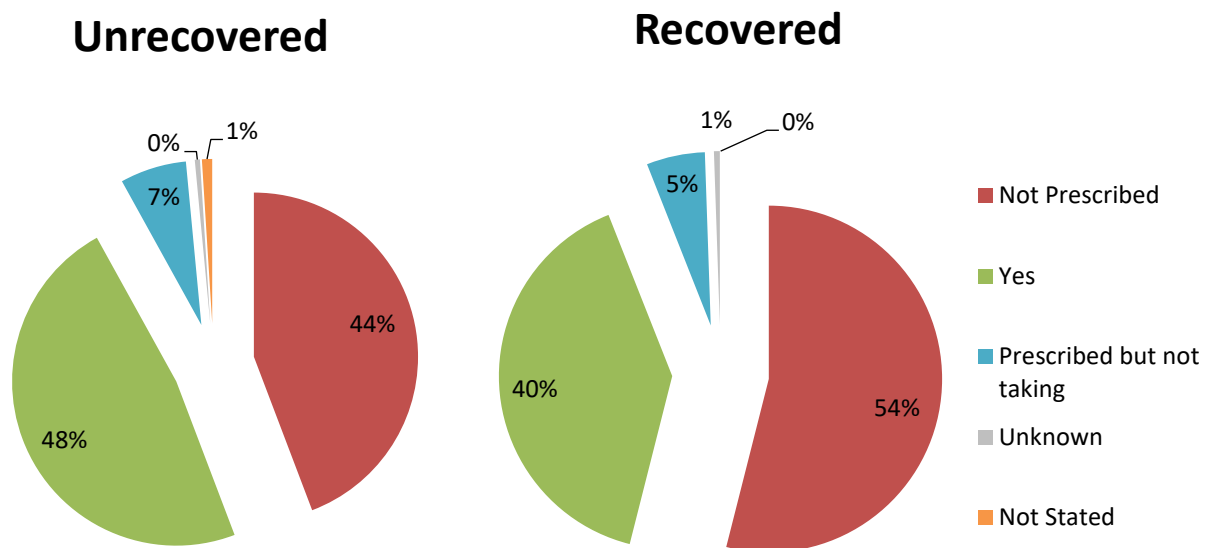
Approximately 10% more clients had not been prescribed medication on discharge in the recovered group compared to the unrecovered group (54.0%,  $N = 198$  and 44.2%,  $N = 88$  respectively). Taking medication was the most frequent response in the unrecovered group, with just under half using this method (47.7%,  $N = 95$ ) compared to the recovered group (40.1%,  $N = 147$ ). There was only 1.1% difference in those who were prescribed but chose not to take their medication, with this being more likely in the unrecovered group (Figure 19).

### ***Are these differences significant?***

A chi-square test of independence was performed to examine the relationship between recovery and medication. The relationship between these variables was not significant ( $\chi^2(2) = 4.46, p = 0.11$ ).

### ***Does the use of medication on discharge predict whether someone recovers or not?***

Within the binary logistic regression model, medication use on discharge was not found to statistically predict recovery ( $b = 0.32, \text{Wald } \chi^2(1) = 1.40, p = 0.24$ ).



**Figure 19: Pie graphs to show the percentage (%) of clients within each recovery group who are prescribed and use medication.**

## 3.2. Service Contact Variables

### 3.2.1. Referral Source

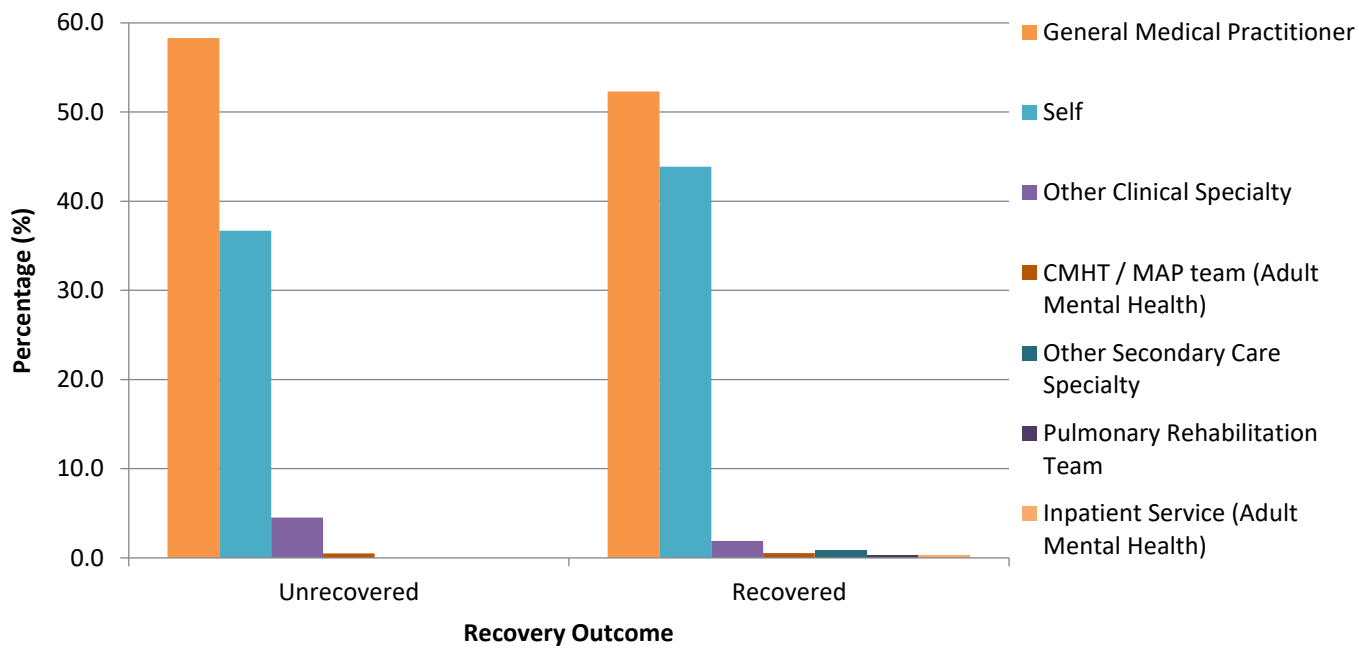
As you can see from Figure 20, there is a greater range of referral sources in the recovered group compared to the unrecovered group. Although for both recovery groups a referral from the GP was the most common, even more so for unrecovered clients (58.3%,  $N = 116$  compared to the recovered group figures of 52.3%,  $N = 192$ ). There was also a greater proportion of self-referrals in the recovered group (43.9%,  $N = 161$ ) compared to the unrecovered group (36.7%,  $N = 73$ ).

#### ***Are these differences significant?***

A chi-square test of independence was performed to examine the relationship between recovery and referral source. The relationship between these variables was not significant ( $X^2(2) = 2.90, p = 0.23$ ).

#### ***Does referral source predict whether someone recovers or not?***

Within the binary logistic regression model, overall referral source was not found to statistically predict recovery ( $Wald \chi^2(2) = 5.01, p = 0.08$ ). However, if the referral was categorised as 'Other' (i.e. not from the GP or self), you were 4.69 times more likely to recover than if it was a GP referral.



**Figure 20: Bar graph to show the type of referral sources and their prevalence within each recovery group.**

### 3.2.2. Number of Previous Referrals

Table 12 presents the statistics for the number of previous referrals to the service for both recovery groups. Those who did not recover from the service had a slightly greater range in referrals, a greater mean number of referrals, and greater spread compared to the recovered group.

**Table 12: The mean number of referrals, with the standard deviations and ranges for each recovery group.**

	<i>M</i>	<i>SD</i>	Range
<i>Unrecovered</i>	0.33	0.64	0-4
<i>Recovered</i>	0.29	0.61	0-3

### ***Are these differences significant?***

An independent T-Test revealed that there was no significant difference between those who did and those who did not recover from this service with regards to the number of previous referrals they had to the service ( $t(564) = 0.64, p = 0.52$ ).

### ***Does the number of previous referrals predict whether someone recovers or not?***

Within the binary logistic regression model, the number of previous referrals was not found to significantly predict recovery ( $b = 0.12$ ,  $\text{Wald } \chi^2(1) = 0.26$ ,  $p = 0.61$ ).

### **3.2.3. Primary Presenting Problem**

Both groups demonstrated a similar pattern in the proportions of primary presenting problems, as in both cases, the most common presenting problem was depression/low mood, followed by anxiety and then panic (Figure 21). These proportions were greater in the unrecovered group compared to the recovered group, except for panic (48.7%,  $N = 97$  vs. 42.0%,  $N = 154$  for depression/low mood; 25.6%,  $N = 51$  vs. 22.1%,  $N = 81$  for anxiety; and 5.5%,  $N = 11$  vs. 6.8%,  $N = 25$  for panic respectively). All other presenting problems were more prevalent in the recovered group, with the exception of: sleep/anxiety, body dysmorphic disorder, low self-esteem and anger.

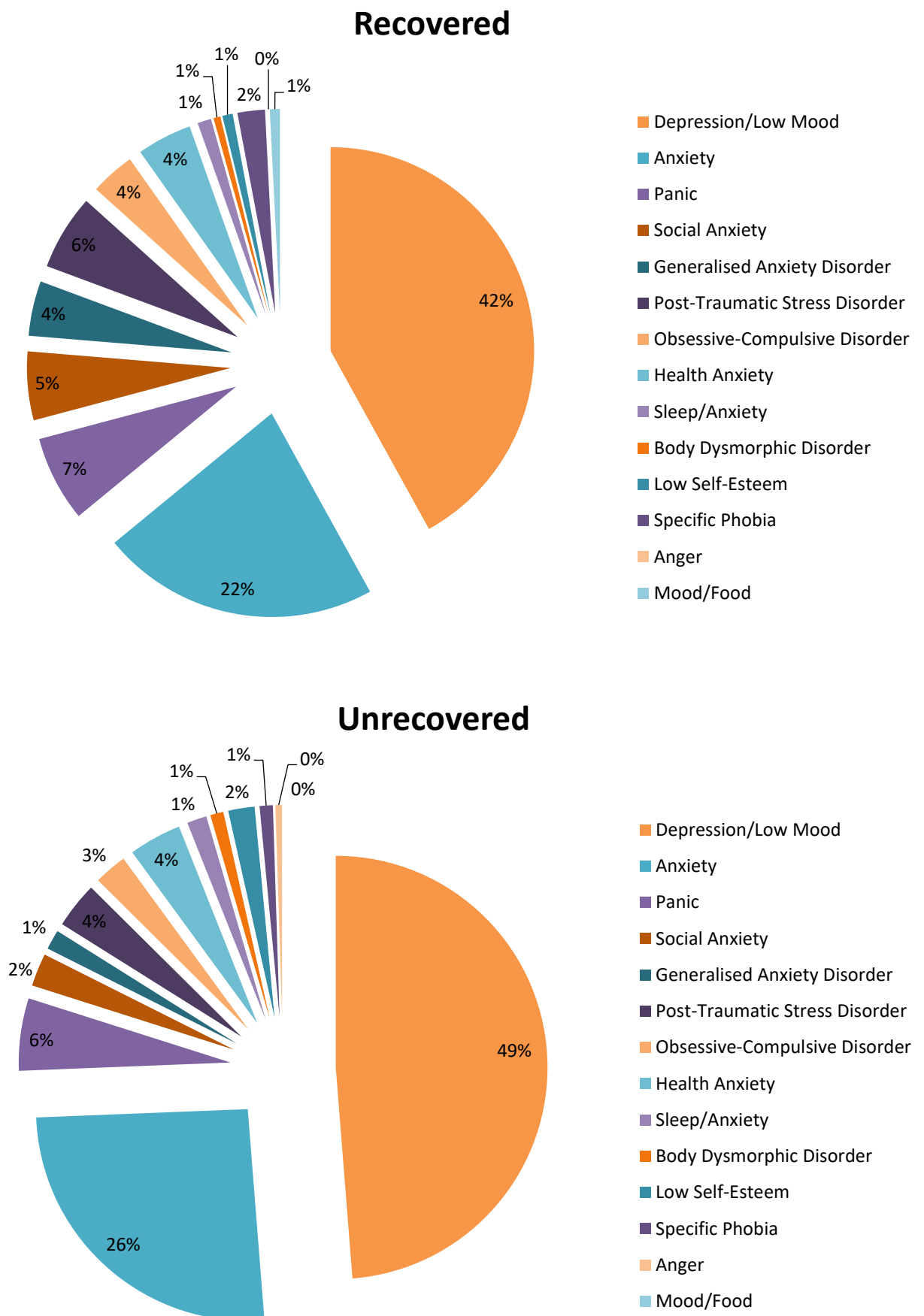
### ***Are these differences significant?***

A chi-square test of independence was performed to examine the relationship between recovery and presenting problem. The relationship between these variables was found to be statistically significant ( $\chi^2(2) = 6.31$ ,  $p = 0.04$ ), therefore there is strong evidence that recovery outcome is associated with presenting problem. In the unrecovered group, there was twice the prevalence of depression/low mood compared to anxiety or 'other' presenting problems (i.e. those not classified as either low mood/depression or anxiety). There were also significantly higher rates of the 'other' presenting problem compared to anxiety in the recovered group.

### ***Does the type of presenting problem predict whether someone recovers or not?***

Within the binary logistic regression model, overall presenting problem was not found to statistically predict recovery ( $\text{Wald } \chi^2(2) = 1.76$ ,  $p = 0.41$ ).

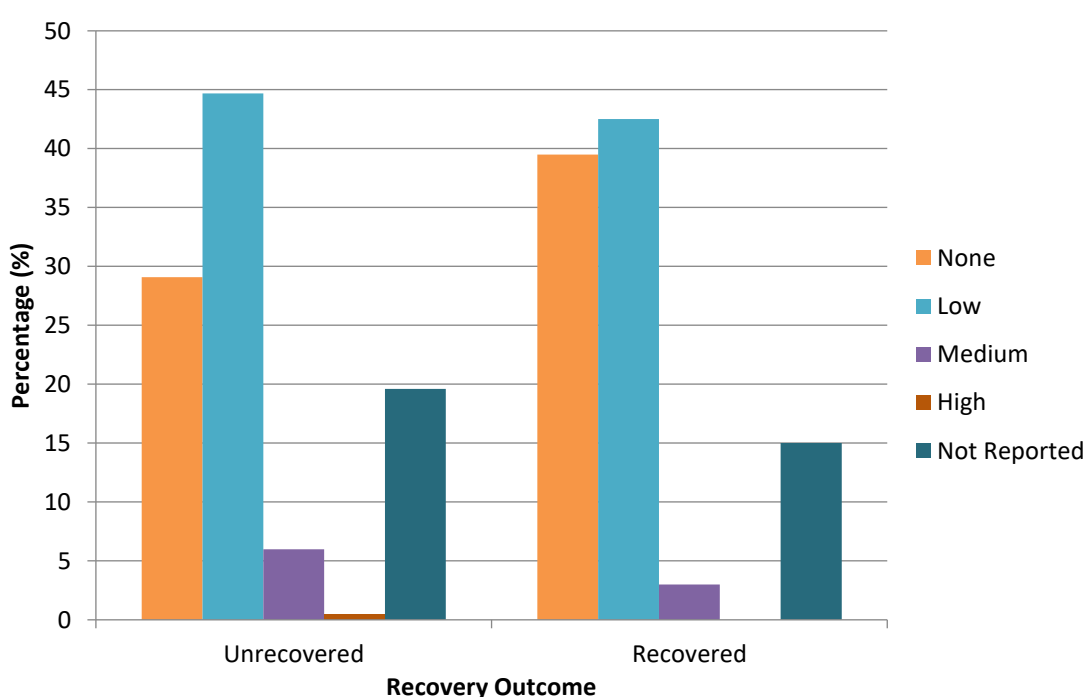




**Figure 21: Pie graphs to show the prevalence (%) of each primary presenting problem within each recovery group.**

### 3.2.4. Risk Level

Clinicians at this service did not report a client's risk level at triage on 34.6% ( $N = 94$ ) of occasions (Figure 22). In the remaining sample, a 'low' risk level was the most prevalent in both recovery groups (unrecovered = 44.7%,  $N = 89$ ; recovered = 42.5%,  $N = 156$ ), although there was over a 10% difference in the number of clients rated as 'no risk'. There were double the proportion of clients considered to be 'Medium' risk in the unrecovered group compared to the recovered group (6.0%,  $N = 12$  vs. 3.0%,  $N = 11$  respectively). There was also one client rated as 'High' risk within the unrecovered group.



**Figure 22: Bar graph to show the percentage of clients within each risk category by recovery outcome.**

To calculate and interpret average risk ratings, a numerical value was given to the various risk categories. Values approximate to 0 represented "no risk"; values approximate to 1 represented "Low"; values approximate to 2 represented "Medium"; and values approximate to 3 represented "High". For the recovered group the mean average risk level was 0.57 ( $SD = 0.56$ ) and for the unrecovered group this was 0.73 ( $SD = 0.62$ ). Therefore, both groups had an average risk level between no-low risk, although it was greater in the unrecovered group.

### ***Are these differences significant?***

With regards to the difference between the average level of risk between the two recovery groups, an independent T-Test showed that this difference was statistically significant ( $t(470) = 2.72, p = 0.01$ ). This suggests that on triage those who do not recover from services are considered to be significantly more risky compared to those who do recover.

### ***Does risk level predict whether someone recovers or not?***

Within the binary logistic regression model, risk was not found to statistically predict recovery ( $b = 0.35, \text{Wald } \chi^2(1) = 1.53, p = 0.22$ ).

#### 3.2.5. Time on Waiting List

Table 13 presents the statistics for the number of days that clients spent on the therapy waiting list following their triage assessment. The unrecovered group waited approximately eight days longer compared to those in the recovered group. The range and variance in the data was also greater in the unrecovered group.

**Table 13: The mean waiting list times (days), standard deviations and ranges for each recovery group.**

	<i>M</i>	<i>SD</i>	Range
<i>Unrecovered</i>	67.50	86.0	0-665
<i>Recovered</i>	59.25	74.51	0-591

### ***Are these differences significant?***

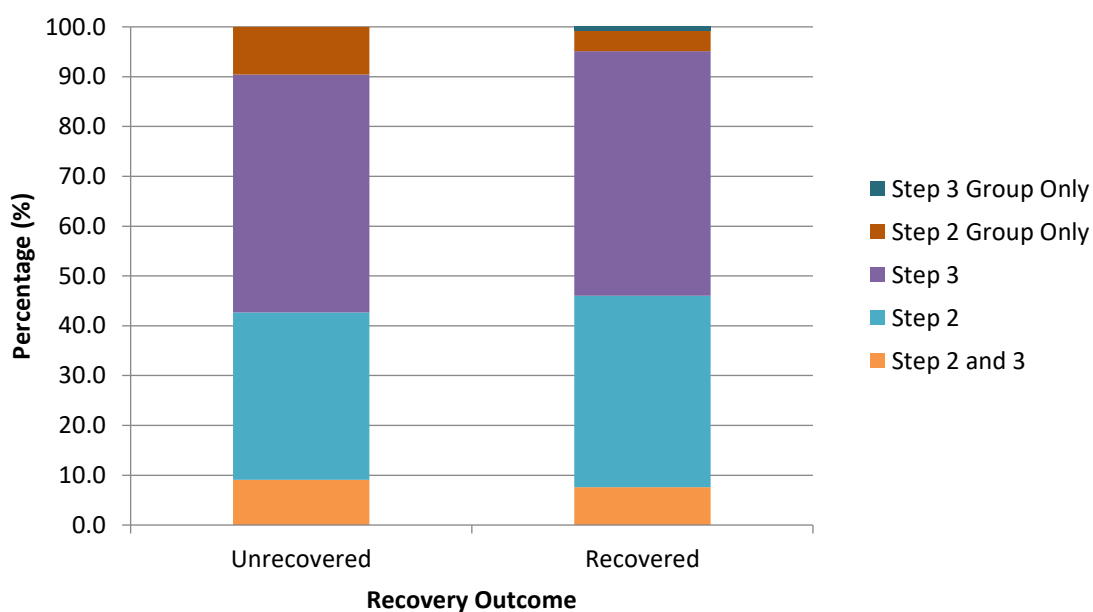
An independent T-Test revealed that there was no significant difference between those who did and those who did not recover from this service with regards to how long they were waiting for treatment ( $t(359.78) = 1.14, p = 0.26$ ).

### ***Does waiting time predict whether someone recovers or not?***

Within the binary logistic regression model, waiting time was not found to statistically predict recovery ( $b = 0.00, \text{Wald } \chi^2(1) = 0.02, p = 0.89$ ).

### 3.2.6. Therapy Type

Just under half of clients were provided with step 3 treatment in both recovery groups, albeit 1.3% greater in the recovered group (49.0%,  $N = 180$  compared to 47.7%,  $N = 95$  in the unrecovered group). Just under 10% of clients in the unrecovered group were provided with group treatment only, whereas this figure was almost half that in the recovered group. Please see Figure 23.



**Figure 23: Bar graph to show the proportion (%) of the different therapy types across recovery groups.**

#### ***Are these differences significant?***

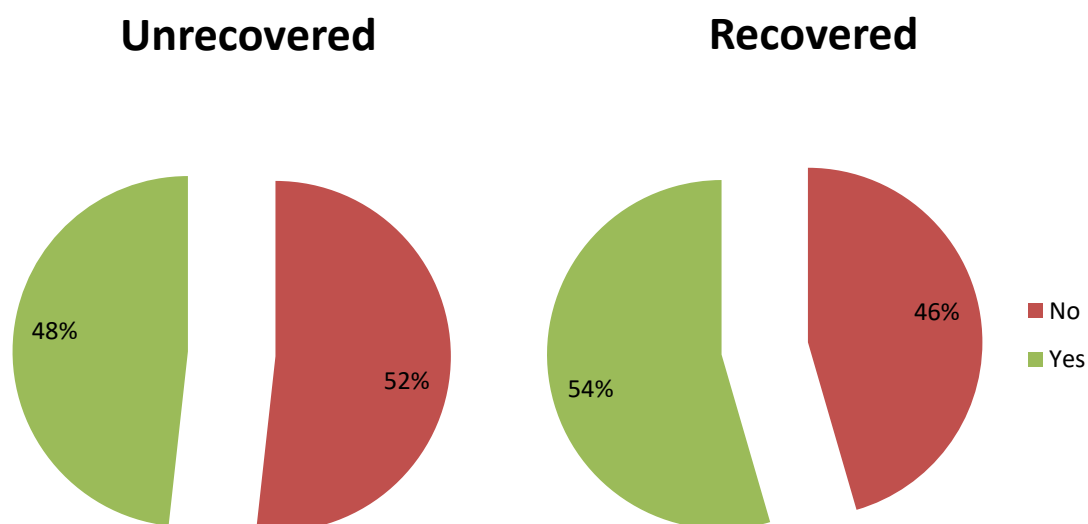
A chi-square test of independence was performed to examine the relationship between recovery and treatment type. The relationship between these variables was not significant ( $\chi^2(3) = 5.41, p = 0.14$ ).

#### ***Does therapy type predict whether someone recovers or not?***

Within the binary logistic regression model, overall therapy type was found to statistically predict recovery ( $\text{Wald } \chi^2(3) = 10.52, p = 0.02$ ). More specifically, compared to group therapy only, clients who received step 2 therapy were 5.23 times more likely to recover ( $p < 0.01, 95\% \text{ CI} = 1.84 - 20.05$ ), and clients who received step 3 therapy were 4.72 times more likely to recover ( $p = 0.01, 95\% \text{ CI} = 1.36 - 16.37$ ).

### 3.2.7. Group Therapy Attendance

There was a difference of 6% regarding those clients who attended a group as part of their treatment (Figure 24). That is, 6% more clients attended a therapy group in the recovered group compared to the unrecovered group.



**Figure 24: Pie graphs to show the percentage (%) of clients attended a therapy group according to their recovery outcome.**

#### ***Are these differences significant?***

A chi-square test of independence was performed to examine the relationship between recovery and group therapy attendance. The relationship between these variables was not significant ( $\chi^2(1) = 2.02, p = 0.16$ ).

#### ***Does attending a group predict whether someone recovers or not?***

Within the binary logistic regression model, attending a group was not found to statistically predict recovery ( $b = 0.08, \text{Wald } \chi^2(1) = 0.72, p = 0.79$ ).

### 3.2.8. Therapeutic Contact Type

As shown in Table 14, both clients who did and did not recover experienced approximately eight individual therapeutic appointments each, although this was slightly greater in the unrecovered group compared to the recovered group (8.1 vs. 7.8 respectively). The unrecovered group

generally had greater mean scores and variance on all contact types compared to the recovered group. Recovered clients had approximately one less appointment booked compared to the unrecovered clients.

**Table 14: The different types of Client-Service contacts along with the mean, standard deviations and ranges for each recovery group.**

Client-Service Contact Type	Recovered			Unrecovered		
	<i>M</i>	<i>SD</i>	Range	<i>M</i>	<i>SD</i>	Range
<i>Client attended appointment</i>	7.8	4.2	0-23	8.1	5.0	1-23
<i>Client DNA appointment</i>	0.4	0.9	0-11	0.6	1.0	0-6
<i>Client was late to appointment</i>	0.3	0.8	0-6	0.5	0.9	0-5
<i>Client cancelled appointment</i>	0.9	1.3	0-6	1.4	1.9	0-10
<i>Service cancelled appointment</i>	0.3	0.6	0-4	0.4	0.7	0-3
<i>TOTAL number of booked appointments</i>	9.9	5.2	1-32	10.9	6.6	1-34

#### ***Are these differences significant?***

The variances of the two recovery groups were significantly unequal for all measures of contact type, therefore a T-test for unequal variances was used. The findings of these are presented below in Table 15. As you can see, there were significant differences found between the recovery groups on the number of DNAs, the number of times a client cancelled an appointment, and subsequently the total number of booked appointments a client had. Within all three circumstances, the unrecovered clients had a significantly higher number of DNAs/cancelled/booked appointments compared to the recovered group.

#### ***Does service contact predict whether someone recovers or not?***

Within the binary logistic regression model service contact was not found to statistically predict recovery: attended appointments ( $b = 0.01$ ,  $Wald \chi^2(1) = 0.01$ ,  $p = 0.71$ ); number of DNAs ( $b = -0.09$ ,  $Wald \chi^2(1) = 0.36$ ,  $p = 0.55$ ); number of times client was late ( $b = -0.13$ ,  $Wald \chi^2(1) = 0.65$ ,  $p = 0.42$ ); number of times the client cancelled ( $b = -0.08$ ,  $Wald \chi^2(1) = 0.64$ ,  $p = 0.43$ ); number of times the service cancelled ( $b = -0.04$ ,  $Wald \chi^2(1) = 0.04$ ,  $p = 0.84$ ).

**Table 15: The results from the independent T-Tests which look at the differences between the recovered and unrecovered group on the number of Client-Service Contacts.**

Client-Service Contact Type	t Value	Degrees of Freedom	p Value
<i>Client attended appointment</i>	0.69	348.05	0.49
<i>Client DNA appointment</i>	2.19	360.25	0.03*
<i>Client was late to appointment</i>	1.34	373.24	0.18
<i>Client cancelled appointment</i>	2.69	294.46	0.01**
<i>Service cancelled appointment</i>	1.30	350.58	0.19
<i>TOTAL number of booked appointments</i>	2.11	336.16	0.05*

\*\* Correlation is significant at the .01 level (2-tailed)

\* Correlation is significant at the .05 level (2-tailed)

### 3.3. Outcome Measure Variables

The findings in Table 16 demonstrate that in both recovery groups, scores on the PHQ-9 and GAD-7 decreased from triage to discharge. On average for both the PHQ-9 and GAD-7, clients in the unrecovered group had greater mean scores on triage and discharge compared to the recovered group. Typically, those who did not recover came into the service with higher outcome scores and showed less change in scores over time compared to those who did recover. Clients also tended to show greater change on the PHQ-9 compared to the GAD-7.

#### ***Are these differences significant?***

Independent T-Tests were used to examine whether the differences between the two recovery groups observed in the table below were statistically significant. The outcome showed that those who did recover from our services had significantly lower PHQ-9 and GAD-7 scores on triage and discharge compared to those who did not recover (Table 17). Individuals in the recovered group also experienced a significantly greater difference in their scores on both the PHQ-9 and GAD-7 compared to the unrecovered group.

Similarly, paired T-Tests were used to determine whether the differences between outcome scores within each recovery group were significantly different over treatment time. Results demonstrated that clients' scores on discharge were significantly less compared to their scores on triage for both measures and in both recovery groups (Table 18).

**Table 16: The means, standard deviations, and classifications of PHQ-9 and GAD-7 scores for each recovery group.**

		Recovered	Classification	Unrecovered	Classification
		<i>M(SD)</i>		<i>M(SD)</i>	
<b>PHQ-9</b>	<i>Triage</i>	13.0 (6.1)	Moderate	16.4 (6.0)	Moderately Severe
	<i>Discharge</i>	3.8 (2.7)	Healthy	12.9 (5.5)	Moderate
	<i>Difference</i>	9.2 (6.2)	-	3.5 (6.1)	-
<b>GAD-7</b>	<i>Triage</i>	11.8 (5.1)	Moderate	14.8 (4.8)	Moderate
	<i>Discharge</i>	3.3 (2.2)	Healthy	12.2 (4.3)	Moderate
	<i>Difference</i>	8.5 (5.3)	-	2.6 (5.1)	-

**Table 17: The results from the independent T-Tests which look at the differences between the recovered and unrecovered group on their PHQ-9 and GAD-7 scores.**

		<i>t Value</i>	Degrees of Freedom	<i>p Value</i>
<b>PHQ-9</b>	<i>Triage</i>	6.33	564	< 0.01**
	<i>Discharge</i>	21.91	252.09	< 0.01**
	<i>Difference</i>	-10.54	564	< 0.01**
<b>GAD-7</b>	<i>Triage</i>	6.68	564	< 0.01**
	<i>Discharge</i>	27.23	258.18	< 0.01**
	<i>Difference</i>	-12.99	564	< 0.01**

\*\* Correlation is significant at the .01 level (2-tailed)

\* Correlation is significant at the .05 level (2-tailed)

**Table 18: The results from the paired T-Tests which look at the differences between a client's score on triage and discharge for the PHQ-9 and GAD-7.**

		<i>t Value</i>	Degrees of Freedom	<i>p Value</i>
<b>Recovered</b>	<i>PHQ-9</i>	28.75	366	< 0.01**
	<i>GAD-7</i>	31.13	366	< 0.01**
<b>Unrecovered</b>	<i>PHQ-9</i>	8.09	198	< 0.01**
	<i>GAD-7</i>	7.29	198	< 0.01**

\*\* Correlation is significant at the .01 level (2-tailed)

\* Correlation is significant at the .05 level (2-tailed)



### ***Do triage scores predict whether someone recovers or not?***

Within the binary logistic regression model, PHQ-9 triage score was not found to statistically predict recovery ( $b = -0.01$  Wald  $\chi^2(1) = 0.01$ ,  $p = 0.94$ ). However, GAD-7 triage score was ( $b = -0.13$ , Wald  $\chi^2(1) = 12.14$ ,  $p < 0.01$ ), where clients with greater GAD-7 triage scores were 0.87 times less likely to recover compared to those who had low GAD-7 scores on triage (95% CI = 0.81 – 0.93).

## **3.4. Recovery Rate**

### 3.4.1. Moving to Recovery

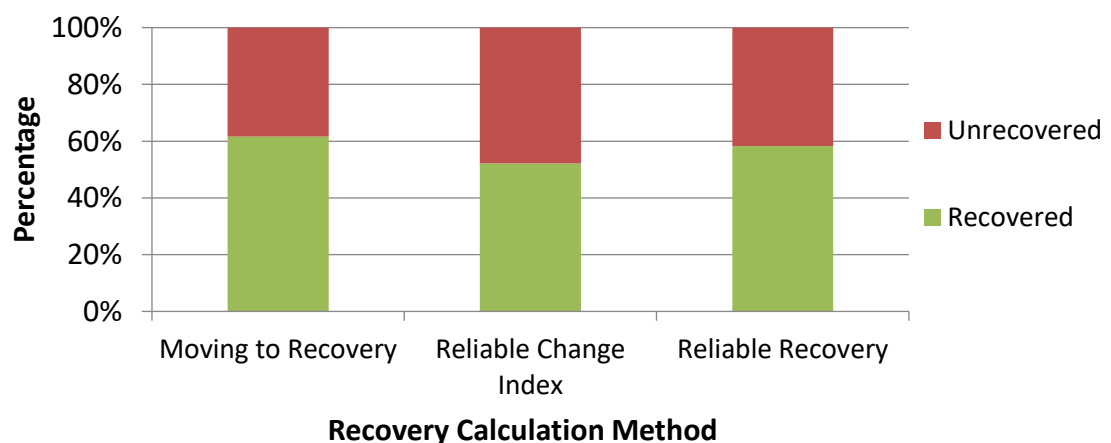
As defined by Griffiths and Steen (2013), 61.6% ( $N = 307$ ) of clients recovered from this service after completing treatment during 2015. Sixty-eight clients were excluded from the calculation due to not meeting the caseness threshold at the start of treatment (total  $N = 498$ ).

### 3.4.2. Reliable Change Index

From the total sample ( $N = 566$ ), a recovery rate of 52.1% was calculated using this method, indicating that 295 clients obtained a statistically reliable improvement in their PHQ-9 and GAD-7 scores, from triage to discharge. This was 12 clients less compared to the “Moving to Recovery” index.

### 3.4.3. Reliable Recovery Index

The proportion of clients who met both the “Moving to Recovery” and “Reliable Change Index” criteria (total  $N = 498$ ), and so can be considered to have reliably improved, was 58.2% ( $N = 290$ ). Please see Figure 25 which presents all three recovery rates.



**Figure 25: Bar graph to show the recovery rates according to each definition.**

## 4. DISCUSSION

---

### 4.1. Summary

This service-related project aimed to identify whether there are any factors associated with recovery and explored whether the current recovery rate calculation was the most useful for this service. The findings demonstrated that there were in fact several variables that differed significantly between those who recovered at this service and those who did not. That is, the unrecovered group were found to include a significantly greater number of clients who: did not speak English; identified as having a long-term condition; were unemployed or reported long-term sickness/disability; primarily presented with low mood/depression; were perceived to be more “risky” on triage; and had more incidences of DNAs and cancellations (and therefore more booked appointments). Moreover, despite all clients demonstrating significant improvements in outcome scores from triage to discharge, the unrecovered group had significantly higher scores on triage and discharge compared to the recovered group, as well as demonstrating significantly less change in scores over treatment. This evaluation also revealed that certain variables predicted recovery at Croydon IAPT, including the type of treatment received (i.e. not group therapy alone); having a specialist referral (i.e. not via self- or GP-referral); being employed; and having a lower GAD-7 score at triage.

These findings are concurrent with previous literature in IAPT services, which also identified similar factors to be associated with recovery. For example, greater symptom severity (i.e. outcome measure scores) has frequently been found to reduce the likelihood of recovery (Gyani et al., 2011; Saunders et al., 2016; Vaillancourt et al., 2015). Although both the PHQ-9 and GAD-7 are usually associated with demonstrating such differences, however here, it was only the GAD-7 measure that predicted recovery which has not been shown to be distinct before. Similarly, being employed (i.e. not receiving public benefit) is a factor identified in the current study as well as in previous research (Saunders et al., 2016), as was the finding that there are no differences in recovery between those who are self- or GP-referred (Gyani et al., 2011), and that recovery was not affected by waiting time (Vaillancourt et al., 2015).

Several of the other findings, however, are inconsistent with published literature, including the effect of the number of appointments a client has. Other studies demonstrate that a greater number of sessions enhance recovery (Gyani et al., 2011; Vaillancourt et al., 2015), whereas the current evaluation showed that there were no differences in the number of therapeutic appointments that recovered and unrecovered clients had. When counting the total number of

booked appointments, the opposite was evident, as more appointments were associated with more DNAs and cancellations and so reduced the likelihood of recovery. It can be argued that more DNAs/client cancellations could reflect less engagement and commitment to psychological therapy, as well as interfering with the consistency of treatment. This would therefore weaken the overall effect of the intervention and make recovery less probable. Similarly, taking medication has been shown to have inconsistent effects on recovery, but this may be explained through the way the variable is measured across studies. In this study, medication use was measured on discharge and shown not to effect recovery, however in Saunders and colleagues' study (2016), medication use was measured on referral, and this was less prevalent in those who recovered. Contrastingly, clients beginning treatment taking psychotropic medication has also been shown to have no impact on recovery after controlling for other confounders (Gyani et al., 2011).

Regarding the most efficient calculation to determine recovery rate, all three methods produced a figure which exceeded the service target of 50%, although it was the "Moving to Recovery" method that yielded the greatest percentage. However, it is important to note that this recovery calculation excludes those clients not meeting caseness at the start of therapy, and therefore does not reflect all client journeys like the "reliable change index" does. Additionally, the sample in the current study was comprised of only those clients who were recorded as completing treatment, which therefore makes it difficult to compare these figures to that of other studies, which would normally include clients who dropped out or were referred on to more appropriate services during their treatment. Not surprisingly then, the current recovery figures are greater than those reported by other studies, which were 41-56% (Chan & Adams, 2014; Clark et al., 2009; DoH, 2012; Gyani et al., 2013; Radhakrishnan et al. 2013; Richards & Borglin, 2011).

However, Atkinson (2014) highlights the difficulties of understanding recovery when multiple methods are available, and how recovery figures can more than half when you additionally consider those clients who, before entering treatment, were referred but exit the service or are referred elsewhere. Therefore, he concludes that the recovery rate statistics are essentially meaningless, as they can be operated to reflect different values according to your agenda. Gyani and colleagues (2011) further observed that those clients with outcome scores in the "severe" range showed significantly greater improvements compared to those with "mild" or "moderate" scores. Yet this was often not considered as "recovery" due to this change not falling below the caseness threshold. Considering this, recovery from the client's point of view may be defined very differently, and so recovery rates may be observed to be greater should, for example, a measure of reaching treatment goals be used, rather than that of clinical symptoms (i.e.

meaningful for the client). Therefore, it is difficult to consider which measure of recovery can be deemed as the most useful for this service, as all limit some client groups to some extent and reporting all methods may become confusing to the reader. Should usefulness be defined by the calculation that places the service in a more favourable light (i.e. the moving to recovery method here), or that is the most inclusive and limits the least number of clients (i.e. the reliable recovery index here), or that is meaningful to the client and considers their perspective? All in all, it is felt that more consideration is required regarding the utility of recovery calculations in general, especially with regards to their purpose, which should aim to place the client's needs first.

## 4.2. Limitations

There are several limitations that need to be considered when interpreting this data. Firstly, this sample reflects only those clients who were recorded as *completing* treatment at the service, which is different to how service performance is usually assessed (i.e. inclusive of all clients who access treatment at the service, including those who do not complete treatment and/or are referred on to other services after beginning treatment). This means that the recovery rates in this project may be an over-estimation compared to what is typically reported to commissioners and senior management. However, again, it raises the question as to how useful this method of reporting is, as it does not reflect the true impact of staff's therapeutic work nor the effects of evidence-based practice.

An observation made during the data collection phase was that approximately 10% of clients in the unrecovered group were those who attended a one-off group (compared to 4% in the recovered group). This may limit and bias the findings as the data from the outcome measures of these individuals would not accurately reflect their recovery. For instance, clients would have been required to complete the measures on triage, and once more on the day they attended the group. However, both scores would reflect times when the client felt psychologically distressed and would not capture a time when they had had a chance to implement the skills learnt, or even "recover".

There are further limits to the data entry and collection in this project. In addition to unavoidable human error in the data collection process, it was also dependent on the efficiency of the staff's data recording. For instance, there was no consistent practice for classifying individuals when they leave the service. The data in the project only consists of those inputted as "completed treatment" on the IAPTus system. However, there could be some misclassification and error in this reporting, for example, some clients could have been coded as "completed treatment"

when the client and therapist agreed to end therapy early and/or were signposted to another service, rather than this being coded as “dropped out” or “referred on”. Additionally, there may have been subjective differences in the primary presenting problem that was recorded, especially as the IAPT services do not formally provide medical diagnoses (as they are a psychological service), and these may have been re-evaluated throughout therapy. Similarly, the outcome measures used by clients to indicate their clinical symptoms are brief measures and based on self-report (PHQ-9/GAD-7). This means that “recovery” in the service is determined by a subjective account, which could therefore be biased and not as reliable as using clinician-rated tools.

### 4.3. Implications and Recommendations

This evaluation was able to highlight several areas that differ between those who do and do not recover from Croydon IAPT, as well as some factors that can predict recovery in clients who complete treatment. The implications from having this knowledge are that the service can consider improvements in their practice which account for these factors and provide more support for those clients. Service users and carers who formed the advisory group for the MAP CAG were consulted about the findings of this service evaluation, as were the current staff team working for the service. In light of these and the project’s limitations, please see Table 19 outlining the recommendations from each group, about how the Croydon IAPT service can promote recovery for its clients. Moreover, it is recommended that further research is carried out into the recovery calculations, to understand their validity and reliability in measuring recovery from both a service and client perspective.

### 4.4. Conclusions

Taken together, the current service-related project has demonstrated that there are significant differences on several demographic and service-related variables between those who do and do not recover from the Croydon IAPT service, some of which also appear to have weight in predicting recovery. The measurement of recovery that can be considered the most flattering to the service is the “Moving to Recovery” method, however, this does exclude some client groups and does not account for more severe cases, which subsequently require even greater change to reach the recovery threshold. Moreover, these findings are only representative of clients who completed treatment, ignoring those who dropped-out or were referred to other services. This raises the question about the utility of such recovery calculations, as they are not

reflective of the actual therapeutic work and evidenced-based practice being undertaken by staff, or the client's perspective of recovery. In the context of these findings, the service has been able to consider several recommendations to promote recovery in Croydon IAPT for future clients and gained an understanding of the effectiveness of their hard work.

**Table 19: Recommendations for key findings as suggested by the ‘Service User and Carers Advisory Group’ and the Croydon IAPT staff team.**

FINDING	RECOMMENDATION	
	<i>Service User and Carer Advisory Group</i>	<i>Croydon IAPT Staff Team</i>
<b>Low mood/depression was the most prevalent disorder in unrecovered group.</b>	None suggested.	To consider an alternative “package of care” for this group. For instance, beginning their care with a behavioural activation group to enhance mood as much as possible before beginning the cognitive therapy. This would also hope to increase motivation to attend and reduce drop out/DNA for this client group.
<b>More clients who do not speak English in unrecovered group.</b>	To have a dedicated language service – training therapists to speak languages and not using interpreters.	To liaise with other local services and share resources, for instance multi-ethnic counsellors.
<b>Significantly higher outcome scores on triage in unrecovered group.</b>	To use and monitor outcome scores to help link and refer on to more appropriate services, should a client’s score on these be too high/low.	To consider using outcome scores as an indicator of what step of care they should receive, and/or referring to an alternative service.
<b>GAD-7 scores significantly predict recovery.</b>	None suggested.	To attend more to levels of anxiety during therapy, even if this is secondary to the presenting problem.
<b>Significantly more DNAs / cancellations / booked appointments in the unrecovered group.</b>	To get consent to contact a next of kin on client’s assessment, so this person can be involved and encourage the client to not drop out/DNA.	To hold to the DNA policy and discharge clients not consistently attending. This would also allow therapists to meet with other clients’ sooner.

FINDING	RECOMMENDATION	
	<i>Service User and Carer Advisory Group</i>	<i>Croydon IAPT Staff Team</i>
<b>Significantly more clients with a Long-Term condition (LTC) in the unrecovered group.</b>	None suggested.	To improve the LTC pathway within the service. People with LTCs are not all being seen within this specific pathway. Staff felt it would be helpful to have more training on LTCs and consider different standards for recovery in this group. For example, it might be acceptable to expect some residual level of anxiety/low mood about the condition. Additionally, some physical symptoms on the outcome measures currently used may be rated by this group according to their physical illness rather than mental health, leading to an overestimation of mental health symptoms.
<b>Data Recording errors (project limitation).</b>	None suggested.	To clarify and define the terms for discharging patients on IAPTus, so all staff are consistently recording data in the same way.
<b>Low recovery in clients attending a one-off group (project limitation).</b>	To provide a follow-up call/session to those attending a one-off group.	To provide a “package of care” for clients attending a one-off group, rather than this one session. This should include a follow-up call to clients, and appropriately discharging them (if recovered), or offering the next step of care.



## 5. DISSEMINATION

---

The dissemination process began by circulating an information sheet to the service's senior management team, highlighting the key findings from the evaluation (see Appendix 8.2). Following this, a meeting with my supervisor and the service manager was arranged, to discuss the initial feedback of the results. The immediate implications of these were considered, as well as clarifying any discrepancies with service performance reports. A plan was subsequently made regarding the further dissemination of the findings, and the write-up of the report.

The next stage of this process began by sharing the results with the 'Service User and Carer Advisory Group' within the MAP CAG. The aim of attending this group was to gather recommendations on how to improve the service, in light of these findings, from a service user and carer perspective. This group was also provided with the 'key findings' handout. Following this, the project was presented at a Croydon IAPT staff team meeting, which was attended by individuals at all staffing levels, including senior management. A PowerPoint presentation was utilised (please see Appendix 8.3 for the handouts), which outlined the aims, methods, and results of the project, as well as the suggested recommendations that were put forward by the 'Service User and Carer Advisory Group'. The staff team who attended this meeting were further encouraged to generate their own suggestions on how to subsequently promote recovery in their service.

Finally, a service evaluation report was written on behalf of the service, incorporating the appropriate recommendations and circulated to the service for final comments. The report also consisted of an action plan which the service agreed to follow in order to implement the recommendations and promote recovery in Croydon IAPT. This was also submitted to the CAG's clinical governance/audit committee for formal monitoring.

## 6. LEADERSHIP

---

This project was an existing idea within the service that I was placed at (as part of my Doctorate in clinical psychology training). Although the senior management team wished to investigate why clients from their service were/were not recovering, they were unsure how to undertake this as a service evaluation project. This allowed me to develop my own ideas about how to gather this information and propose them to my supervisor. I developed a preliminary plan and made the contacts/connections to relevant bodies to understand the feasibility of the project. For instance, I contacted the IAPTus team to identify whether the data I required could actually be pulled from the electronic patient system and arranged for my account to be approved to access this function. Once the senior management team were happy with the proposed project's aims and methods, I went on to set up the relevant meetings and complete the appropriate paperwork to obtain approval from the Clinical Governance Team and involve service users as much as possible in this planning stage. By taking the lead and developing the service's idea into a practical project with little guidance, it helped me to appreciate the importance of good decision-making and time-management, as well as building on my organisation and communication skills.

Moreover, several minor problems arose whilst undertaking the project, such as the sheer quantity of the data being collected which was proving difficult to manage by one person in the allocated time. I was therefore required to discuss this with the University and consider all the options of taking this forward. The outcome was to recruit another member of staff to assist with this the project. This further enhanced my leadership skills as I was then required to motivate this person to engage them with the task, as well as manage and organise their involvement. Towards the final stages of the project, I organised the dissemination of the findings, adapting the method to suit relevant audiences. Regarding the implementation of change, I further had to attempt to motivate the service team to engage in such discussions, as well as action plan what this would look like. This was achieved through presenting the project to the team, emphasising the positive findings relating to their work (e.g. achieving above the 50% recovery rate target), and encouraging them to come up with their own recommendations about how the service could promote recovery, rather than these being enforced by someone who does not have as much experience of working there. I finally closed my involvement in the project by providing a complete handover to my supervisor regarding the expectations following the project (e.g. implementing the action plan and monitoring this).

## 7. REFERENCES

---

- Atkinson, P. (2014). Lies, Damned Lies, and IAPT Statistics. *Self & Society*, 42, 18-19.
- Chan, S. W. Y., & Adams, M. (2014). Service Use, Drop-Out Rate and Clinical Outcomes: A Comparison Between High and Low Intensity Treatments in an IAPT Service. *Behavioural and Cognitive Psychotherapy*, 42, 747–759.
- Clark, D. & Oates, M. (2014). *Measuring Improvement and Recovery in Adult Services: Version 2*. Retrieved from <http://www.oxfordahsn.org/wp-content/uploads/2015/11/measuring-recovery-2014.pdf>.
- Clark, D. M., Layard, R., Smithies, R., Richards, D. A., Suckling, R., & Wrighte, B. (2009). Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. *Behaviour Research and Therapy*, 47, 910–920.
- Department of Health. (2011<sub>a</sub>). *Talking Therapies: A four-year Plan of Action*. Retrieved from [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/213765/dh\\_123985.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213765/dh_123985.pdf).
- Department of Health. (2011<sub>b</sub>). *No Health Without Mental Health*. Retrieved from [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/213761/dh\\_124058.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213761/dh_124058.pdf).
- Department of Health. (2012). *IAPT Three-year report: The first million patients*. Retrieved from <http://www.mhpf.org.uk/sites/default/files/documents/publications/iapt-3-year-report.pdf>.
- Griffiths, C. A., & Griffiths, L. J. (2015). Recovery and reliable change rates for patients scoring severe on depression, anxiety or impaired functioning in a psychological therapies service: IAPT. *Mental Health Review Journal*, 20, 28-35.
- Griffiths, S. & Steen, S. (2013). Improving Access to Psychological Therapies (IAPT) Programme: Setting Key Performance Indicators in a More Robust Context: A New Perspective. *The Journal of Psychological Therapies in Primary Care*, 2, 133–141.
- Gyani, A., Shafran, R., Layard, R. & Clark, D. M. (2011). Enhancing Recovery Rates in IAPT Services: Lessons from analysis of the Year One data. Retrieved from <http://www.iapt.nhs.uk/silo/files/enhancing-recovery-rates--iapt-year-one-report.pdf>

- Gyani, A., Shafran, R., Layard, R. & Clark, D. M. (2013). Enhancing recovery rates: Lessons from year one of IAPT. *Behaviour Research and Therapy*, 51, 597-606.
- Health and Social Care Information Centre. (2015). *Psychological Therapies: Annual Report on the use of IAPT services England, 2014/15*. Retrieved from <http://www.hscic.gov.uk/catalogue/PUB19098/psyc-ther-ann-rep-2014-15.pdf>.
- IAPT National Programme Team (2011). *The IAPT Data Handbook (Version 2.0)*. Retrieved from <http://www.iapt.nhs.uk/silo/files/the-iapt-datahandbook.pdf>.
- IAPT Programme. (2016). *IAPT: Improving Access to Psychological Therapies*. Retrieved May 22, 2016, from <http://www.iapt.nhs.uk/>.
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The Phq-9. *Journal of general internal medicine*, 16, 606-613.
- National IAPT Programme Team. (2012). *IAPT Key Performance Indicator (KPI) Technical Guidance for Adult IAPT Services 2012/13 (Version 2.0)*. Retrieved from [https://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjpp7f1wszQAhULBMAKHTxDD3AQFggbMAA&url=http%3A%2F%2Fwww.wirral.org.uk%2Findex.php%2Fdownload\\_file%2Fview%2F255%2F307%2F&usg=AFQjCNEZLIU-GQzGU5rCGsPPFO5FI-WAqQ&sig2=oR5IUBDPk9gK7LF5wwK7dg&bvm=bv.139782543,d.ZGg](https://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjpp7f1wszQAhULBMAKHTxDD3AQFggbMAA&url=http%3A%2F%2Fwww.wirral.org.uk%2Findex.php%2Fdownload_file%2Fview%2F255%2F307%2F&usg=AFQjCNEZLIU-GQzGU5rCGsPPFO5FI-WAqQ&sig2=oR5IUBDPk9gK7LF5wwK7dg&bvm=bv.139782543,d.ZGg).
- Radhakrishnan, M., Hammond, G., Jones, P. B., Watson, A., McMillan-Shields, F. & Lafortune, L. (2013). Cost of Improving Access to Psychological Therapies (IAPT) programme: An analysis of cost of session, treatment and recovery in selected Primary Care Trusts in the East of England region. *Behaviour Research and Therapy*, 51, 37-45.
- Richards, D. A. & Borglin, G. (2011). Implementation of psychological therapies for anxiety and depression in routine practice: Two year prospective cohort study. *Journal of Affective Disorders*, 133, 51–60.
- Saunders, R., Cape, J., Fearon, P., & Pilling, S. (2016). Predicting treatment outcome in psychological treatment services by identifying latent profiles of patients. *Journal of Affective Disorders*, 197, 107–115.
- Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine*, 166, 1092-1097.

Vaillancourt, K., Manley, J., & McNulty, N. (2015). Why has our recovery rate dropped? An audit examining waiting times, starting scores and length of treatment in relation to recovery within an IAPT service. *The Cognitive Behaviour Therapist*, 8, 1-14.

## 8. APPENDICES

### 8.1. PHQ-9 and GAD-7 Questionnaires

#### PHQ- 9

Over the <u>last 2 weeks</u> , how often have you been bothered by any of the following problems?	Not at all	Several days	More than half the days	Nearly every day
1 Little interest or pleasure in doing things	0	1	2	3
2 Feeling down, depressed, or hopeless	0	1	2	3
3 Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4 Feeling tired or having little energy	0	1	2	3
5 Poor appetite or overeating	0	1	2	3
6 Feeling bad about yourself — or that you are a failure or have let yourself or your family down	0	1	2	3
7 Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
8 Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3
9 Thoughts that you would be better off dead or of hurting yourself in some way	0	1	2	3

#### GAD-7

Over the <u>last 2 weeks</u> , how often have you been bothered by any of the following problems?	Not at all	Several days	More than half the days	Nearly every day
1 Feeling nervous, anxious or on edge	0	1	2	3
2 Not being able to stop or control worrying	0	1	2	3
3 Worrying too much about different things	0	1	2	3
4 Trouble relaxing	0	1	2	3
5 Being so restless that it is hard to sit still	0	1	2	3
6 Becoming easily annoyed or irritable	0	1	2	3
7 Feeling afraid as if something awful might happen	0	1	2	3

## 8.2. 'Key Findings' hand-out

### **Key Findings**

- The Recovery Rate for Croydon IAPT in 2015 = 61.6%. This is greater than the target of 50%.
- Clients scores on discharge were significantly less compared to their scores on triage for both measures (PHQ-9 and GAD-7) and in both recovery groups ('Recovered' vs. 'Unrecovered') – i.e. on average all patients showed symptom improvement.

### **Differences found between those who 'recover' and those who do not:**

- There were significantly more clients who do not speak English in the unrecovered group compared to the recovered Group.
- There were significantly more clients who recover when they do not have a LTC compared to when they do.
- In the unrecovered group, there were significantly fewer clients who were employed, and significantly more clients who were unemployed and who were long-term sick/disabled.
- When comparing the proportions of primary presenting problems in each group (categorised as: (1) low mood/depression, (2) anxiety, or (3) other specific disorders), the recovered group had a significantly smaller proportion of anxiety compared to 'other' (i.e. there were more 'specific' disorder diagnoses compared to general anxiety).
- The unrecovered group had a significantly greater proportion of clients with low mood/depression compared to 'other' (i.e. there was more low/mood depression compared to 'specific' disorders).
- On triage the unrecovered group were considered to be significantly more risky compared to the recovered group.
- Unrecovered clients had a significantly higher number of DNAs/cancelled/booked appointments compared to the recovered group.

- The recovered group had significantly lower PHQ-9 and GAD-7 scores on triage and discharge (i.e. less severe symptoms) compared to the unrecovered group.
- Individuals in the recovered group also experienced a significantly greater difference in their scores on both the PHQ-9 and GAD-7 (i.e. a greater improvements in symptoms) compared to the unrecovered group.

Recovered Group	Unrecovered Group
Less non-English speaking clients	More non-English speaking clients
More clients without a LTC	More clients with a LTC
More employment Less unemployment Less long-term sick/disabled	Less employment More unemployment More long-term sick/disabled
More likely to have 'other' as presenting problem compared to anxiety. LM/depression still the most prevalent.	Twice as likely to not recover if have LM/depression compared to anxiety or 'other' (which were equal).
Less risky on triage	More risky on triage
Less DNAs/cancellations/booked appointments	More DNAs/cancellations/booked appointments
Lower PHQ-9 and GAD-7 scores on triage and discharge	Higher PHQ-9 and GAD-7 scores on triage and discharge
Greater difference in PHQ-9/GAD-7 scores from triage to discharge	Smaller difference in PHQ-9/GAD-7 scores from triage to discharge

**Factors found to predict recovery:**

- Clients who were unemployed were 0.36 times less likely to recover compared to those who were employed.
- Clients considered as Long-term sick/Disabled were 0.34 times less likely to recover compared to those who were employed.
- Clients whose referrals were categorised as 'Other' (i.e. not from the GP or self), were 4.69 times more likely to recover compared to if it was a GP referral.



- Clients who received step 2 therapy were 5.23 times more likely to recover compared to Clients who received group therapy only.
- Clients who received step 3 therapy were 4.72 times more likely to recover compared to Clients who received group therapy only
- Clients with greater GAD-7 scores on triage are 0.87 times less likely to recover compared to those who have low GAD-7 scores.



## 8.3. PowerPoint presentation for service staff



### Recovery in IAPT

- “Caseness”
  - When a client is considered a clinical case and not yet recovered (i.e. their symptoms are sufficiently severe to be considered a clinical problem).
- Determined by the weekly outcome measures
  - PHQ-9: Score of 10 or above = Caseness threshold
  - GAD-7: Score of 8 or above = Caseness threshold
- Figures are used by management and commissioners to evaluate service performance
  - Target = 50% recovery

### Types of Recovery

- “Moving to Recovery”
  - Proportion of clients who move from clinical caseness at the beginning of therapy to below caseness threshold by discharge
  - What IAPT uses - \*\*Required on BOTH the PHQ-9 & GAD-7\*\*
- “Reliable Change Index”
  - Considers whether the change in scores pre/post treatment can be considered statistically reliable
- “Reliable Recovery”
  - When both the “Moving to Recovery” and “Reliable Change Index” criteria have been met.

### What Have I Been Doing?

- Primary Aim:
  - To investigate the factors associated with recovery in Croydon IAPT
  - Why might clients not recover from our service?
- Secondary Aim:
  - To investigate whether the current recovery calculation is the most useful for Croydon IAPT
  - What is the most reflective measure of recovery?

## What Have I Been Doing?

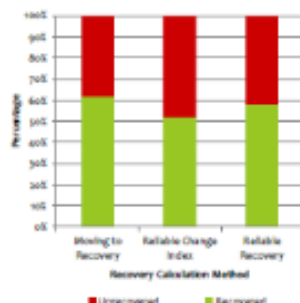
- Sample:
  - All Clients who were discharged after completing therapy
  - Between 1<sup>st</sup> January 2015 – 31<sup>st</sup> December 2015
- 566 clients identified from IAPTus
  - 65% = “Recovered”
  - 35% = “Unrecovered”

## What Have I Been Doing?

Demographic Information	Service Contact	Outcome Measures
- Gender	- Referral Type	- PHQ-9 score on Triage
- Age on discharge	- Number of Previous Referrals	- PHQ-9 score on Discharge
- Ethnicity	- Primary Presenting Problem	- Difference in PHQ-9 score from triage to discharge
- Sexuality	- Risk Level at Triage	- GAD-7 score on Triage
- Language (English)	- Time spent on the Waiting List	- GAD-7 score on Discharge
- Disability	- Type of Individual Therapy Received	- Difference in GAD-7 score from Triage to Discharge
- Long-Term Condition	- Attendance to a Group	
- Employment on Discharge	- Number of attended individual therapeutic contacts	
- Medication use on Discharge	- Number of times the client 'Did Not Attend' (DNA)	
	- Number of times the client was late	
	- Number of times the client cancelled	
	- Number of times the service cancelled	
	- Total number of booked appointments	

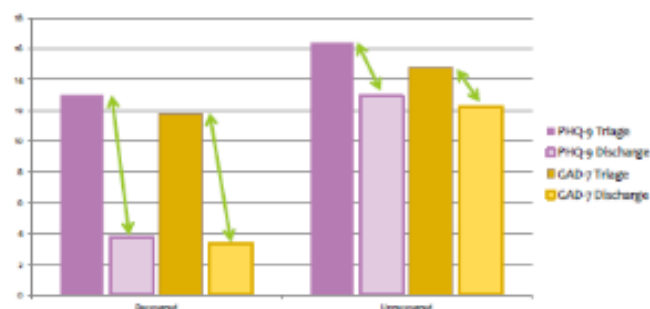
## Key Findings

- “Moving to Recovery”
  - 61.6%
- “Reliable Change Index”
  - 52.1%
- “Reliable Recovery”
  - 58.2%



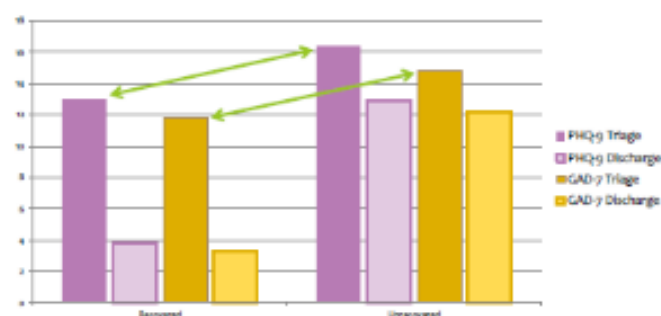
## Key Findings

- Overall, clients' scores on discharge were significantly less compared to their scores at triage for both groups.



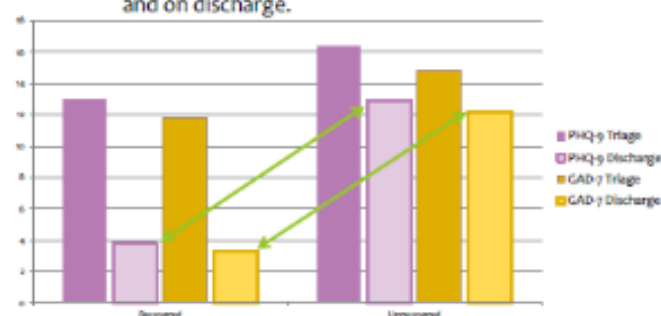
## Key Findings

- Clients in the 'Recovered' group had significantly lower scores on triage...



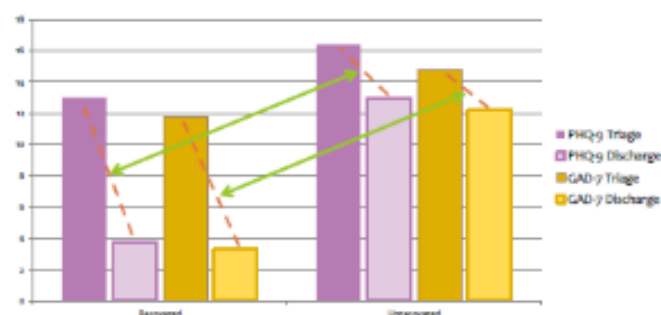
## Key Findings

- Clients in the 'Recovered' group had significantly lower scores on triage... and on discharge.



## Key Findings

- And those in the 'Recovered' group also had significantly greater difference in their triage and discharge scores.



## Key Findings: Other sig. Differences

Recovered Group	Unrecovered Group
Less non-English speaking clients	More non-English speaking clients
More clients without a LTC	More clients with a LTC
More employment	Less employment
Less unemployment	More unemployment
Less long-term sick/disabled	More long-term sick/disabled
More likely to have 'other' as presenting problem compared to anxiety, LM/depression still the most prevalent.	Twice as likely to not recover if have LM/depression compared to anxiety or 'other' (which were equal).
Less risky on triage	More risky on triage
Less DNAs/cancellations/booked appointments	More DNAs/cancellations/booked appointments

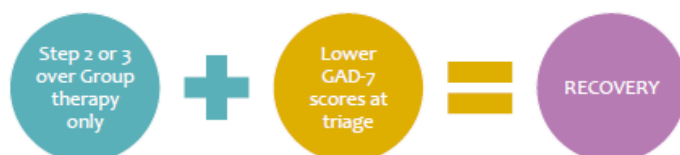
## Key Findings: Predictors of Recovery

- More likely to recover if...



## Key Findings: Predictors of Recovery

- More likely to recover if...



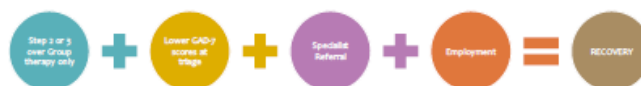
## Key Findings: Predictors of Recovery

- More likely to recover if...



## Key Findings: Predictors of Recovery

- More likely to recover if...



## Things to Consider...

- Only those who completed treatment
  - Ignores those who dropped out/referred on
  - Underestimation of true recovery rate?
  - But is it useful to include these groups?
- Recovery negatively influenced by those who attend a one-off group
  - No opportunity to apply skills and to see change
  - No follow-up

## Recommendations??

- MAP CAG's 'Service User and Carer Advisory Group':
  - Provide a follow-up call/session to those attending a one-off group
  - Having a dedicated language service – training therapists to speak languages and not using interpreters.
  - Using and monitoring PHQ-9/GAD-7 scores to help link and refer on to more appropriate services.
  - On admission get consent to contact next of kin to involve them should the client drop out/DNA.



## Recommendations??



## Acknowledgements

- And a **HUGE** thank you to Jane Ellis and Damian for helping me with this project!!

**THANK  
YOU  
FOR BEING  
AWESOME**